Horizon 2020
INFRADEV-1-2014 - Design studies

RICHFIELDS Working Packages 8
Deliverable D8.1

Report on case studies

Date delivered:
M15

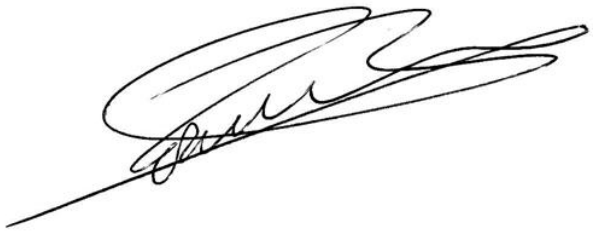
Author:
Bent Egberg Mikkelsen, Kwabena Titi Ofei, Haris Hondo, Erik Kaunisto

Deliverable lead beneficiaries:
AAU

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 case studies
Deliverable No.:	D 8.1
Authors:	Bent Egberg Mikkelsen, Kwabena Titi Ofei,Haris Hondo, Erik Kaunisto
Reviewer:	Karin Zimmermann – Project Coordinator prof. dr. ir. Pieter van't Veer – Scientific Coordinator
Start date:	M6
Delivery date:	M12
Due date of deliverable:	M15
Dissemination level:	Public
Status:	

Change history:		
Version	Notes	Date
001		11-01-2017



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.

Contents

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

Executive summary

As part of the overall RICHFIELDS project aimed at designing a future European research infrastructure (RI) for innovative research on healthy food choices, preparation and consumption of EU-citizens the WP8 has carried out research on that topic. The present study aims to investigate opportunities, problems and challenges of business generated consumer data created by data providers and stakeholders in both the private and public sectors. The study aims to discuss the potentials of such data as input a future RI. For this purpose, interviews were conducted in Denmark and Sweden – the 2 countries selected as the empirical foundation of WP8. The interviews were carried out with representative from Gothenburg City (GC), Inköpscentralernas aktiebolag (ICA), Statistics Denmark (SD) and Gesellschaft für Konsumforschung (GfK) and the idea was to use cases representing both business-to-business, B2C, business to consumer and business to government, B2G. The interviews aimed at examining the potentials for extracting information of scientific significance for the RI, examining suitable user group strategies to the data platform for different stakeholders and exploring possible access strategies including aspects of rights and responsibilities. The interviews were thematic analysed. The content for each interview was coded according to relevance. From the combination of codes, themes were extracted and harmonized to understand the underlying meaning of the information content. The joint themes emerging from the analysis were: company description, data generation purpose, data management and structure, data analysis, data usage, data access authorization, collaborative data sharing, data quality control and consent and ethical issues.

With respect to the scientific case it could be concluded that diversity in data sources is considered important for the RI in order for the RI to be able to provide answers to broad socio-economic and consumption behaviour related research questions. However, differences in data generation purposes may put limits on data relevance, which must be considered by the RI. Incomplete, missing or erroneous data must also be considered. This would require access to raw data, knowledge of EU-legislation in terms of data storage times and RI data integrity and scrutiny measures, respectively. Further if the use of individual's identifiable data is relevant for scientific purposes, the RI potentially needs a related EU-legislation for intra-EU consumer identification.

Regarding user group strategies, it can be expected that the stakeholders of the public data providers may also be interested in the services of the RI. In addition, it could be concluded that public sector data providers as compared to market research business seem more likely to be willing to share their data freely and potentially learn more about how to work more pro-actively to support researcher's work on health issues. On the other hand, with respect to the private sector it is important that the RI can ensure mutual benefit and add value for data providers in addition to what they are already able to do themselves. This can be done by e.g. exchanging more complete data sets in exchange for raw data. Otherwise, payment for data is another solution.

Finally, in terms of access strategies, the RI can expect a varying degree of access authorization from its data providers. Without an overall EU-legislation around

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

simplification of data sharing with the RI, specific terms and conditions may have to be negotiated with each data provider. Careful handling of sensitive information in terms of raw data is also emphasized, and the RI is likely to require specific EU policies on data handling, data aggregation and disclosure to meet the concerns of data providers.

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

Introduction

There is growing interest in consumer health related to food, behaviour and lifestyle determinants. At the same time digitalisation of societies create new digital patterns and sets of data on consumer behaviour – for instance in the cases of food choice and consumption. These data sets generate new potentials, potentially leading to new methods of analysing population behaviour. However, data is fragmented, key information is lacking, and the resulting knowledge gap prohibits policy makers and companies to make effective strategies regarding public health nutrition and for enterprises to reformulate food products and to develop new strategies which increases the availability of healthier foods.

In particular, two segments of consumers are interesting when it comes to big food data: the domestic consumption which is a proxy for individual food and health behaviour and the institutional food consumption that over the past decades have been significantly increasing. This type of business-to-government (B2G) trade between wholesalers and caterers is a good proxy for the type of food consumption and food choice that is made in the institutions of the public-for instance in schools, kindergartens, hospitals etc. Procurement of healthier and sustainable foods for the public plate has become an important target for European policy makers because it is speculated to be able to contribute to more sustainable food consumption. [1] [2] [3] [4]. The importance of public expenditure on buying of goods is undisputable [5] [6] [7] and is regulated by a number of policies and rules at both national and international level. Policy makers have learnt about the different ways through which the food economy can be influenced in a more sustainable and healthy way through the implementation of new practices of procurement. The value of food bought by the public for public sector settings such as hospitals, schools, universities, care homes, armed forces and canteens is considerable [8] [9] [10] and foods in the public sector represent a significant part of the food economy in many countries [11].

The digital consumer society seems to be offering scientists new methods, which can increase the understanding of purchasing habits and food choices. And it seems to be creating the opportunity for a future data sharing economy where consumers are donating their data open up new avenues for food business analysts, marketing researchers and scientists. In addition to consumer generated data also business collects and analyse huge amounts of data.

Already government agencies and business are generating and sharing their data with scientists for instance interested in measuring the effect to public health interventions and natural experiments. Phenomena such as effect of removing VAT from fruit & veggies (FV) [12], effects of FV nudges in supermarket and limiting intake of sugar-sweetened beverages (SBB) has been studied using business generated data [13,14]. However, the full potentials of using individual dataset might not be unlocked until researchers and scientists are able to access and use combined dataset from different sources from a common platform. A common platform for business and organisation to share their individual generated data will facilitate the availability of the combined data for researchers to address new research questions. This forms the basis of the Research Infrastructure on Consumer Health and Food Intake using E-science with Linked Data Sharing (RICHFIELDS) Project. This is seeking to

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

develop state-of-the–act research infrastructure that combines knowledge of consumer behaviour and food intake in one data platform. RICHFIELDS is a H2020 funded program with 17 European organisations aiming at designing a future research infrastructure (RI) for innovative research on healthy food choice, preparation and consumption of EU-citizens, closely linked to their behaviour and lifestyle. This RI is scheduled to bridge the gap by linking the agricultural-food and nutrition-health domains and account for the regional and socio-economic diversity of the EU. The RI will be instrumental to produce a scientifically reliable, technically sound and socio-legally robust evidence-base that enables scientists to efficiently collect, unlock, connect and share research data of EU-citizens. The consumer-focus and the scientific evidence of RICHFIELDS will, through its services, be available to (a) EU consumers and consumer platforms, (b) stakeholders along the food chain, and (c) policy actors in the agricultural-food and nutrition-health domain.

As a part of the overall RICHFIELDS this current study aimed to investigate what could be the best practices in cases where buying and procurement behaviour can be extracted from existing business generated data and the potential opportunities and challenges of linking such data to the RICHFIELDS platform .The findings of the study are believed to have important implications for the RI and would help to address related concerns including governance of data, privacy issues and ethics as well as the implications for business.

Methods

Study design

This study adopted an exploratory qualitative approach in order to gain an in-depth understanding of what constitutes best practices of generated purchasing and procurement data from informant perspective within their organization. Multiple cases were used in this study. The four cases were represented by four experts familiar with their respective organisation or institutional generated data. This case study approach was considered appropriate as it allowed us to explore the complexity and diversity of stakeholders in a real-world setting [11].

Selection of cases

The selection of the four cases was done through purposive sampling. The rationale behind this was to obtain diversity of responses relevant for identifying best practice. In total four cases, two from Denmark: Gesellschaft für Konsumforschung, GfK and Statistics Denmark (SD) and two from Sweden: Göteborgs Stad, (City of Gothenburg)(GC) and Inköpscentralernas aktiebolag, (ICA), were selected by two experts whom are partners of the RICHFIELDS consortium. This selection criterion was based on the combined prior knowledge of WP 8 leader and the Swedish partner regarding these organisations' purchasing/procurement generated data strategy and their different IC technologies and devices employed to capture and store the data. These case organizations among others contacted were the only ones that expressed an interest in the project and were willing to provide potential data for the analysis. In terms of size of purchasing/procurement data coverage, for instance Statistic Denmark is the main organization collecting B2G organic procurement data. In the case of GfK, it one of the largest retail market research organisation in Denmark and has the highest B2C purchasing market share In this report we

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

understand best practices as cases that align with WP8 selection criteria, cases that are generally accepted as well renowned entities in the sector and cases that are willing to serve as respondents for the RICHFIELDS research.

Ethics

Prior to the interview all the informants were briefed about the research purpose, assured them of confidentiality of any information provided and considered to be contentious to their organisation. All the informants gave an oral consent to be interviewed.

Data Collection

From each of the selected cases, one expert was interviewed leading to four interviews that were conducted. We defined an expert as one employed by the case organisation and that has insight in the data generation and extraction, structure and use of the data. The interview data was collected using a semi-structured interview guideline. This guideline was developed based on insight from the literature and WP8 partners' input including the discussions on the Copenhagen phase 2 kick off in March 2016. The interview guide was structured to cover questions related to:

- purpose of generating the data
- data structure
- relevance of data content
- challenge(s) of using the data to the benefit of the retail industry
- data needs and how they can be addressed
- opinions about using generated data to address societal challenges
- relevance/ constraints of sharing data with the research community
- privacy policy and ethical issues

The semi-structured nature of the interview facilitated flexibility to follow up on new information presented in the context of the interview, and explores new findings in depth. The interviews were piloted in both Denmark and Sweden. The feedback from the interviews was used to clarify the questions. The interviews were conducted at the work place of the participants and audio recorded after permission has been sought. Each interview lasted approximately 80 minutes.

Desk Research

In order to provide a detailed analysis, the two interviewers from Sweden and Denmark requested and obtained a sample of a purchasing/procurement document of each case organization. It is important to emphasize that the document research was to identify the structure, content of the dataset rather than focusing on the analysis of data itself.

Data Analysis

We followed the six thematic analysis phases proposed by Braun and Clarke [16]. These phases are: (1) Data familiarization (2) Codes generation (3) Themes search (4) Themes review (5) Themes definition (6) Themes reporting. Two researchers from each country (DK and SE) performed the thematic analysis. One researcher handled the two interviews conducted in Denmark and another researcher handled the two interviews conducted in Sweden. The researchers who conducted the interviews were both PhD and MSc holders,

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

but the analysis was performed in a consistent manner only by PhD holders, thus well aware of the applied methodology. The group also have experience from using current methodology in previous studies. To familiarize with the data, all four interviews were transcribed and read at least twice in the beginning. The researchers then coded the transcripts independently by assigning relevant names to highlight relevance of text, using QSR Nvivo 11. The coded results of the two researchers were compared and disagreements were discussed to improve clarity. The two researchers worked on codes and collated codes in potential themes independently. The researchers jointly discussed each theme thoroughly looking for similarities and differences. The contents were agreed and those that generated disagreements between the researchers were discussed further until consensus was reached. Finally, the two researchers re-read the data and assigned experts to the themes to their respective data and followed by crosschecking to enhance accuracy.

Results

This section of the report presents the results of the four case study interviews analysed under the following key themes.

Company description

GC

Gothenburg City is a public procurement company located in Gothenburg, Sweden. The organization is very large and enables and handles public procurement of large quantity of foodstuff and associated contract agreements for the district administrations, i.e. various city functions and areas such as the city theatre, pre-schools, schools, assisted living and elder care. For the purpose of extracting “Business to Government” associated information (B2G) for WP8 in the RICHFIELDS project, Anita Olofsson Enquist from Gothenburg City was therefore interviewed. Anita works in the procurement company as a “meal controller”, implying follow-up of e.g. contracts and procurement related statistics/data.

ICA

Inköpscentralernas aktiebolag, ICA is a Swedish food and health retail company. The company is relatively large with an approximate 50% market share and uses consumer generated data to continuously evaluate, adapt and improve their market strategy. For the purpose of extracting “Business to Consumer” and “Business to Business”-associated information (B2C and B2B) for WP8 in the RICHFIELDS project, Joel Ringbo from ICA was therefore interviewed. Joel works as a consumer analyst at ICA, where their transaction database is used in order to analyse consumer purchase data and find patterns.

GfK

Gesellschaft für Konsumforschung, GfK is an international research market organisation with operations in three Nordic countries, Denmark, Sweden and Finland. GfK Denmark deals directly with country specific market research activities but also collaborates with the other Nordic partners on diverse tasks such as data-sharing. GfK is one of the key organizations conducting market research in the Danish food retail sector. They are well noted for providing data and consulting services to the players in the food retail sector on consumer buying behaviour and preferences. By virtue of this GfK’s operation, it can be regarded as “Business to Consumer”-organisation. In order to get an insight into the generated data of

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

the organization, Kenneth Johansen, Nordic Manager in-charge of Dictionary & Coding, Operations, Nordics Shopper and Consumer Panel was interviewed.

SD

Statistics Denmark, SD is a state-owned institution mandated by Law to collect, compile and publish statistics on the Danish Society. SD uses national register such as business register of active enterprises (ESR) to contact registered businesses and request for any data of interest. The data suppliers are mandated by law to provide the data requested by SD. This institution aims to produce national statistics that are accurate and reliable, coherent and comparable. The organizational structure is divided into five sections including the business statistics. Within the business statistics, the division for the food industries is responsible for generating national data on procurement/sales of organic food and beverages in retail trade. On this basis SD functions as "Government to Business" organisation provider of organic food procurement data to the diverse group of users including researchers and students. Martin Lundø, chief Adviser to the Food Industries Division was our key informant.

Key information for Richfields Research platform

The table 1 below contains themes for understanding of the value for each potential actor collaboration with the Richfields Research platform. This provides key information to feed into the development of Richfields Research platform and possible business models.ble

Summary of themes relating to case organisations

Themes	GC	ICA	GfK	SD
1. Types of data that the organization currently holds	Supplier provided procurement data	Business generated consumer data	Consumer grocery purchasing data	Organic food sales/procurement data
2.Customer base (users of data)	Media Journalists Agencies GC (due to political decisions)	ICA Researchers Suppliers	Retail business Researchers	Policy makers Researchers/Students Government agencies Retail and food service businesses
3. Data needs and perceived benefit (s) to the organisation	Only public procurement data from district administrations with varying degree of aggregation This data provides regional coverage of public procurement behaviour. The data may be useful to probe	The data may possibly not include consumer loyalty data due to privacy legislation issues. ICA has a 50% market share and such organizations can provide valuable insights in national/regional consumer purchasing behaviour	Household members' food and beverages purchase at convenience shops or market are not captured. There exist vacuum in the food retail industry, capturing this data will be useful to gain	Organic food sale data from the foodservice is not captured under the current data collection. The collection of this data will increase coverage and provide better understanding of organic food consumption at the consumer level.

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

	differences in public procurement due to different policies.		new insights and understanding of consumer purchasing behaviour in convenient market.	
4. Existing collaboration with other RIs or other commercial organisations for data acquisition/data sharing	No specific collaborations. Stakeholders are mostly media, journalists and governmental agencies.	ICA likes research and has collaboration with Stockholm School of Economics and Örebro and Jönköping University. The research is motivated by finding patterns in consumer behaviour.	GfK collaborates with Euro panel. It is a non-financial collaborations which permits GfK to share data on Euro panel global platform.	SD currently has no collaboration with RI and other commercial organisations for data acquisition and data sharing.
5. Types of data willing to be shared with the Richfields platform?	Data sharing seems to be feasible. The data may however have different levels of aggregation.	Data or partial data sharing seems to be feasible under certain conditions due to ICAs stock market exposure.	Sharing of aggregated data with Richfield seems feasible under special terms as done in the case of Euro panel. Generally, GfK sells aggregated data to non-commercial research institutions and retail businesses.	SD will prefer to share aggregated sales/procurement data with Richfields platform. SD's aggregated data can be shared and accessed free of charge.
6. What types of data they would not be willing to share with the Richfields platform?	No specific limitations have been noted. However, the data may possibly be aggregated to limit regional resolution.	ICA cannot share information compromising consumer privacy, i.e. loyalty data. The data is therefore likely to be anonymized and possibly aggregated.	GfK may reluctant to share raw data containing individual identifiable information. Sharing of this data will require approval from individuals donating the data.	Raw/micro data is not shared by SD because it contains enterprises sensitive data which needs to be protected.
7. Perceived benefits for collaborating with the potential Richfield's platform.	GC are likely to be interested in contributing to public health and may also get valuable input on	ICA is motivated by increased consumer understanding in order to maximize profitability and	Data sharing with researchers aimed at addressing societal related problems could	Data sharing with RICHFIELDS was perceived to be a good opportunity for SD to understand how it can be used to address

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

	their own data management.	consumer satisfaction. Simultaneous contributions to consumer health research are considered positive.	be basis for non-financial collaboration with RICHFIELDS.	societal health issues. In addition, SD can identify future data needs of researchers and incorporated it into subsequent surveys.
8. Perceived risks/barriers for collaborating with the potential Richfields platform.	GC have expressed no specific risks. A possible barrier may however be a tight schedule in relation to required amount of involvement.	ICA realizes that their data may be valuable to competitors if potentially leaked from the Richfields platform. Mutual benefit conditions between ICA and Richfields may have to be negotiated.	GfK benefits financially by selling data to research institutions and anticipates the risk of losing revenue if their prospective clients can assess the data free of charge.	Collaboration will be a challenge because of the obstacle to deliver micro data.

Data Generation Purpose

GC

The procurement process at Gothenburg City is a result of a balancing act with respect to their suppliers and clients as well as the interests of political stakeholders. In order to meet political decisions, the organization needs to collect statistics/data from their suppliers to be able to follow development and analyse the procurement process. For this purpose, Gothenburg City has both a "Strategy for Food Group" and a "Contract Group" to facilitate procurement. The "Strategy for Food Group" has an important role in the procurement planning process, where earlier procurements are analysed and goods are matched, but also in ensuring political decisions on e.g. procurement of organic food. Established agreements are usually of framework type, spanning over a four-year period and handled by the "Contract Group". The contracts are continuously modified or updated with regard to specific products and monitored one to three times a year. The agreements include demands and criteria for the suppliers to pay a commission and provide statistics/data on their products with an associated deadline. Failure to meet the deadline results in a supplier penalty, this done in order to cover resulting organization costs. On the other hand, the continuous and future availability of suppliers is considered an important issue.

"Well, yes, if there are no suppliers you cannot order anything. The consequence is that there is no agreement and it is bad. Yesterday me and my colleague met in the "Strategy for Food Group" where we discuss how we should think in the next procurement. We have a "Contract Group" that we meet tomorrow. We meet every month and discuss, for example, if we are to have this tomato or not, or a specific product. There are various functions in the city."

ICA

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

As a large retail company in a competitive business, ICA has developed both a dynamic market and consumer focus, which is dependent on handling and analysing business-generated consumer data. On the market side, the central organization uses data to provide assortment recommendations to the respective ICA store-owners, but since they are free to choose their own assortment, there may be local variations. The performance of competitors is also closely monitored with buying associated consumer-generated data from the "Nielsen Company", followed by appropriate strategic actions. On the other hand, ICA also sell their own consumer generated data to their suppliers and other organizations and has a tradition of sharing their data with universities for mutual benefit research purposes. On the consumer side, ICA focusses on analysing their consumer base by segmentation in order to find consumer patterns.

"ICA don´t want to be the "Big Brother" telling people what they should do or not should do. We just want to provide products for our consumers."

GfK

Gfk as market organization generates household panel data to fill gap in market knowledge related to consumer groceries purchasing behaviour in the Danish retail sector. GfK household panel survey data collection is all year round. The collected data is analysed to create value for understanding of customers' purchasing-behaviour such as who they are, where they shop and what they buy.

"So if you worked for Danish Supermarket and you know everything purchased at Danish Supermarket. You don't know what they bought in Coop and in the other chains. I know that you buy some of your groceries at Danish Supermarket and some of them somewhere else and I can actually help you get the client to move the purchases from them to you".

SD

SD generates organic procurement data for monitoring levels of organic consumption in the Danish society and policy formulation. Apart from SD being the main user of the generated data other stakeholders from food industries including the Ministry of Environment and Food, Danish Agriculture and Food Council and Organic food interest depend on SD data source. Researchers and students equally rely on the data for their projects.

"You are talking about the purpose of collecting this data number and one of the purposes is to fulfil benchmarking of growth of organic food as continuation of previous government and the other purpose is for Statistics Denmark to have a make complete picture of organic consumed".

Data Management and Structure

GC

Gothenburg City collects data from their suppliers twice a year and the data is then further distributed to the various city areas and neighbouring municipalities. The data management is influenced by political decisions thus the organization has regular telephone meetings with their suppliers with feedback on the provided statistics/data. The data collection

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

process makes no demands on the business systems of the suppliers, but is associated with challenges regarding how to adapt deadlines when the suppliers need to provide the statistics/data. Current deadlines are in June and December and if the data is submitted later, the penalty is issued. Gothenburg City often needs to spend extra time to give feedback to new suppliers and those who are not consistent when they submit their statistics/data. On the other hand, at the same time it has been recognized that the general understanding of the data collection process has increased among the suppliers. It is however difficult for the organization to know if there are errors in the supplied data. This is also further complicated by the level of data aggregation, which generally depends on the type and size of the contract. In the contract for the wholesaler, the amount of data aggregation is high. This makes it impossible to see in detail which purchases are made and what, since otherwise, the data files would be too large for the system to handle.

"I have kept with me what I have received over a 4-year period. We didn't put such demands and quality 4 years ago as compared today. We did not get data from every supplier, or received data too late so that we could not follow up, for example, the purchase of coffee. We have corrected this now, because it's interesting for us to know before the procurements."

ICA

In order to facilitate and improve data analysis at ICA, the data management is structured around both collecting the relevant raw data and addressing associated challenges. The raw data generated in the stores consists of date of purchase, bar code, item type, which store, quantity, cost, time of day and whether the purchase was made online. Furthermore, if a loyalty card is used, the consumer ID is also registered. The data is sent to the ICA IT department, where additional data on item category and whether it is sold per kg or piece is added to each transaction. The consumer generated data is stored for a maximum of 18 months in a database together with bought data from "Nielsen Company" on competitors and the data is available to the business analysts in both raw and aggregated forms. Regarding data collection challenges, ICA recognizes both consumer and store related issues. It is important to get all consumers to use their loyalty ID in order to be able to efficiently track their consumption behaviour. This can be achieved by enabling the possibility to e.g. register a bank card/credit card, which would also help to better the identification of certain product categories in small purchases, where loyalty ID often is not used. Central ICA has also noticed that many stores have missing or incorrect information on their local assortment, which is partly attributed to the use of non-standard, so called PLU-codes. However, attempts have been made to standardize these codes in collaboration with GS-1 in Sweden.

"Yes, it is a database text file. It's like the information you brought with you. We have many different tables then and one table might look like this. Then we have a table with all bar codes. Then we have information which category it belongs, if it is private label or not, ecological or not, brand. Then we can match those two tables and get information about this or that. We have several different tables that we can connect to each other. So, yes, basically it looks like this, or in a text file. It is digital. It is in a database."

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

GfK

The GfK household panel data structure is built on four pillars. The shop, covers information about the shop chains; household, includes information on age, income, family size and its location; the article identifies what is in the shops, product barcode, type of product; and finally the movements depicting actual purchases of the households. All these four are put together as a household food panel for the analysis of the purchasing behaviour. Regarding to data collection challenges, the generated data does not cover individual household members purchase at the convenience shops or market. This is partly due to the fact that the member in charge of the household shopping is the same person responsible for scanning of the purchased groceries. This makes it impossible for other household members without the scanning device to scan their purchase outside the routine household purchases. This creates gap in the entire household purchasing data captured. The GfK is aware of this challenge and also the growing demand for data on consumer purchases at convenient market.

"*It would absolutely be better, but I don't think we will ever get the full picture. Because we are looking at household purchases, and a household consists of two, three, four, five people who all have their needs and they all go the 7eleven shop and buy their own soft drink and not all are participants in this panel".*

SD

Organic food product sales data is generated annually by SD. Statistics Denmark obtains information from major supermarket chains and wholesalers selling food to retail shops and foodservices. Organic food sales data provided by these enterprises serve as an indicator of organic food composition levels in the retail and the foodservices. The generated data is structured as turnover, product category and customer groups. The organic food sales turnover is distributed over a number of commodities and measured in weight and values. The product category shows the various food products sold and the customer groups describes the type of retailer or foodservice specific, procuring the organic food products. SD current data capturing does not offer option for the retail and foodservice businesses to report their sales data. Collecting this data in addition to the current data capture will increase the coverage and provide better understanding of the pattern of organic sales and consumption.

"*The foodservice sector is a missing link in our statistic because today the greater part of what we eat, take place in a restaurant, canteen and institution and for that reason there is need to know how big that part of the issue".*

Data Analysis

GC

The data/statistics provided by the suppliers according to the contract is reported into specific software. The software "WINST" is then used to analyse the costs and volumes by comparing the supplier reported sales with what was actually bought in order to detect possible deviations. However, the suppliers have nothing to gain from reporting e.g. lower commission, as this can be easily checked and results in future non-procured goods of the concerned type. In certain cases, specific purchases can be traced in "WINST" if the

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

procured volumes are not too big. “WINST” is also used to search for specific product information, which is based on supplier provided product sheets that can be accessed through the database software “DABAS”. Furthermore, in relation to political decisions, Gothenburg City uses “WINST” to follow the development of organic food on an organizational level. The information is also broken down and sent to the respective city functions and municipalities for them to track development more in detail.

“You can follow up the operations in Gothenburg in a program called "WINST" that handles electronic orders. The reporting tool can show for example a school or a nursing home and see exactly how it looks there.”

ICA

Consumer generated data is essentially used to find patterns among consumers, identify trends and forecast consumer behaviour in relation to assortment changes. The quality of the analysis is strongly connected to consumer ID availability. However, in cases where joint analysis is performed with the separate ICA bank, consumer ID’s, if available, cannot be shared. The output of the consumer analysis forms a basis for strategic decisions within ICA and includes e.g. prediction of effects of discount campaigns, prediction of holiday consumption and associated product supply issues in stores and at central ICA and analysis of important items. In addition, analysis is not only performed at central ICA, but the store owners themselves also have an interest in knowing how well they are doing in order to take appropriate strategic actions. The purpose is to continuously develop the assortment in the right direction, both at central ICA and in the stores.

“As an analyst we can use this information for so much, e.g. to develop our assortment, find the best and most preferred products, what are the trends? If a product isn’t popular, people tend to buy it more seldom. Then we can delete it from the assortment. It costs us so much to have it in the assortment so we can delete it from the system.”

GfK

One important finding highlighted in the interview was that GfK analysis and provides aggregated data to deduce purchasing decisions and behaviour patterns of consumers. Analysis that could provide insights into consumer unhealthy product purchases behaviour seems not be of interest to GfK. GfK perceives this as an attempt to influence household purchasing behaviour. It was revealed that analysing the data to provide information on consumer unhealthy choices could influence type of purchase to report and bias the data.

“They don’t get it now days, like 5-8 years ago when they got in tables of data so that they could analyse it themselves. But today we present it, we actually want to tell them what to do with their business, help them with their business problem instead of just giving those data as we did in the old days”.

SD

SD performs analysis on data collected from enterprises and presents it in aggregated form. The analysed data provides insights into organic food consumption patterns. It was obvious that the analyses does not cover social and health related aspect of organic procurement. This could be due to the fact that SD focuses on generating sales dataset, thereby limiting

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

the possibility of analysing the data for social and health implications of organic food procurements.

"And then again the foodservice sector in general we do not have much knowledge about this. This is a quite huge sector both in employment, in terms of money and in terms of influence on people's life".

Data Usage

GC

The data/statistics collected by Gothenburg City is used both for procurement and to answer specific questions from different stakeholders. In the procurement related data analysis, it is important to track volumes, costs and development of e.g. organic products, but also to check that the suppliers have paid the commission. Regarding stakeholders, questions may come from various sources, such as the environmental agency, journalists and researchers. While journalists may wonder about ethical issues in procurement, such as the origin of the chicken, researchers may connect the statistics to health issues. On the other hand, the amount of detail is limited in the data, since it is aggregated, and data is currently available four years back in time. Researchers may however influence the way the organization works with data by providing feedback to the data management.

"We also use statistics / data before purchasing to see the history. We use statistics when you come and ask. The environmental administration agency is interested in it and we deliver statistics / data there. Journalists can wonder about something specific and then we can look at any specific question."

ICA

ICA uses the consumer data for different purposes, e.g. answering simple bonus related consumer questions, aggregation and data selling (mainly to suppliers), development and evaluation of targeted discount campaigns and prediction of holiday consumption. However, the usability of the data is strongly connected to the quality and output from the data analysis and is investigated through joint research efforts with e.g. Stockholm School of Economics, Jönköping University and Örebro University, where ICA supplies data with the aim of increasing the understanding of consumer behaviour. ICA appreciates research and with a large market share of about 50% they have the ability to estimate overall trends in consumer behaviour.

"We have collaboration with Stockholm School of Economics in Stockholm. They do a lot of research and we share our information with them. We make the information anonymous. It is not possible to identify consumers. We like this kind of activities and we are open-minded. Stockholm School of Economics doesn't have to pay us. We have collaboration with them. We give them information and they give us information back. There is a guy here at ICA that has been a PhD student together with ICA at Örebro University. I also know a guy who was sitting down in Jönköping."

GfK

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

Retailers and researchers mainly use GfK data. Retailers purchase analysed data for understanding purchasing behaviour of consumers such as what they are buying and where they buy their groceries. The data driven market information purchased by the retailers enable them to grow their businesses. On the other hand non-commercial research institutions, such as universities and PhD students purchase data from GfK based on their research needs.

"I am pretty sure that the researchers at Copenhagen University are much better than we are. They probably know our data much better than we do."

SD

According to SD, the data is used by interest groups and authorities including the ministry of food and research institutions to assess the level and composition of organic consumption and related research questions. SD publishes aggregated data on organic procurement for businesses, general public and research institutions. SD is also in regular contact with stakeholders to ensure that the User Committee for Food statistics data needs are covered.

"Yes absolutely. It can also be in the preparation of the surveys that we contact the researchers to say if I am using the concept the right way. But not only researchers, but I also contact industry associations. Of course there are also many experts in the municipalities. This is very important for the statistics. And also to get feedback on the use of data."

Data Access Authorisation

GC

Gothenburg City is able to provide statistics/data upon request from stakeholders. Information on procured volumes, product labels and nutritional values are accessible for e.g. researchers and media. There is no direct data access for stakeholders. Instead, data access authorisation is granted certain people within the organisation that, in turn, can provide the necessary information. The time frame for data access is currently four years back in time.

"You must have access authorization. In each administration and company there is a person called "vinst-administrator" or similar title. This person gives all the other people in his neighbourhood or administration access. Everyone is not authorised to order in "vinst", for example me. There are some people among us who can though. It is determined by each company or administration which persons have these different permissions."

GfK

Currently retail businesses and research institutions pay to access aspects of the generated data. The data, which can be accessed by these organisations, to some extent, is anonymized. GfK does not grant researchers and businesses direct access to their data platform, rather the data is extracted for them. This enables GfK to maintain control over data that can be accessed by retailers or researchers. Non-commercial research institutions requesting access to the extracted data needs to satisfy the organisations' requirements and agree to the terms and conditions stated in the agreement before the data can be accessed.

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

"*Copenhagen University gets every movement and every household with all the details, so if they wanted to perhaps they can actually start digging and find people yes. We have to extract it from our own system and give it to them. They don't have access directly. I prefer that we have things separate."*

SD

SD grants access to anonymise micro data for research purposes. SD has laid down rules governing access to anonymize micro data. It specifies who can get access to the data and what data can be accessed. The user group defined includes all employees in government funded research projects, public institutions such as the universities, government research institutions, ministries etc. The user group from private sector includes non-governmental organizations, consulting firms and other individual enterprises with permanent research and analysis environment in Denmark. Foreign researchers may be granted access to micro data provided they are affiliated with authorized Danish institution, which can assume the overall responsibility of the data access. Researchers granted authorisation normally access the data from SD servers. It is also possible to transfer data to the researchers' server in that case remote access is provided through the internet. It is a requirement that the researcher signs an agreement with SD and agreed on the terms and conditions before access can be granted. This restricts the user or researcher from identifying enterprises, persons or removes any micro data. Researchers may be permitted to remove only aggregated data with no possibility to identify enterprise from the server.

"It is very difficult to make data confidential, and to anonymize these data because the concentration of the enterprises. So this is one obstacle to deliver microdata, the other is that there is actually not much to come for in the microdata. I cannot make data much richer than I am already doing. There is not much more in the micro data".

Collaborative Data Sharing

ICA

Due to stock market exposure, sharing data for research purposes may be a sensitive issue, but becomes less complicated if only partial data is considered. Specific projects have to be discussed within the organization, where also possible regulations have to be considered.

"One problem with sharing is that we are introduced to the stock market now, which means that the security around the information is much higher now compared to before. I guess that is a problem with sharing data with others, since you can track how it goes for a company and maybe speculate on the stock market. I'm not sure about that, but I think there are others at ICA and those are more into exactly what we can do and can't do. Our standpoint however is positive, but there might be regulations that we have to follow, of which I'm not currently sure of."

GfK

GfK believes that sharing data with researchers to address behaviour and lifestyles challenges could be beneficial to society. In view of this GfK has demonstrated its readiness to collaborate with RICHFIELDS on data sharing for research. At the EU level, GfK collaborates with Euro panel on data sharing. This is non-financial collaborations which

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

permits GfK to share data on Euro panel global platform. This is a clear indication that there exist opportunities for non-financial collaboration that the RICHFIELDS consortium can explore to its advantage.

"Research companies that have panel data in the whole EU and actually do a project these days called BG20. We deliver every year to them, data from every country, and they are able to aggregate and do research."

SD

Data sharing on research infrastructure was perceived to be a positive opportunity for the institution. This can help SD to have deeper insight of how the existing generated data could be used to address societal and health issues. In addition to this, SD will be in the position to identify future data needs of researchers and incorporated it into subsequent surveys.

"*It is a general strategy that we should be open to the research community and their needs. So this is why I enter this corporation to get more knowledge about the data I use. We have our survey now, but maybe it should be changed in two, three years, maybe we should make a new survey"*

Data Quality Control

GC

Current data content limitations in the supplier provided statistics/data make it impossible for Gothenburg City to directly trace the origin or manufacturing country of a product. However, through the supplier provided product sheets in the database "DABAS", which is accessible through "VINST", it is possible to trace the name of the company behind a certain product. In this way, the organization may sometimes indirectly figure out where a product comes from.

"There is something called "DABAS" that you can look into. The statistics / data says not everything, for example, if a product is Swedish or not, but you can see when you read the title which company it comes from."

ICA

Missing or erroneous data in the consumer generated data are significant problems at ICA. Due to the large assortment and associated high amounts of data, it is essentially impossible for ICA to trace errors in the product information. Furthermore, certain products that are often part of small purchases, measuring of consumer loyalty might often be impaired, since a loyalty card is often not used in these cases. This implies difficulties for ICA to trace the performance of certain product categories. On the other hand, consumer identification is achieved in approximately 75% of all sales and facilitated use of the ICA card together with efficient discount campaigns are identified as tools to obtain better data quality.

"One thing, as I explained, was the quality problem, i.e. that we don´t have information on all products. That is a big obstacle. Another is that all consumers don´t use their loyalty card, which makes it difficult for us to track e.g. which products should be removed. It's an

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

obstacle for us when all consumers don´t use their loyalty card to be able to identify their purchase."

GfK

GfK performs rigorous quality check to remove erroneous information inputted by the household panellist. The demographic information is compared to the national database and adjusts for any discrepancy. For any suspicious information like extremely high or low purchases reported, the panellist is contacted for confirmation. The information check is done internally before the data is loaded onto the local reporting solution.

"Yes, often they just register two beers and a flower week by week. Nobody can live with that. Perhaps they eat out all the time, so we call them and have a discussion with them about what they register, and say: "listen up, you just have registered two beers and a flower, how do you live?".

SD

In the case of SD all questionnaire survey information inputted by the respondents are extracted from the database and put into an error correction system, using SAS. The system is used to detect any suspicious data for corrective action.

"Let's say, if there is double reported turn over on one commodity then there is a high probability of error, but is it actually true? If they have a change in their composition of goods, we will call them and check this, and we also check the kilo price of the goods. This is the data error correction."

Consent and Ethical Issues

GC

The ethical issues that Gothenburg City works with are mainly related to their demands on the supplier provided statistics/data. It is important that ethical procurement choices can be made among the different suppliers. However, at the same time, the organization also strives to be reasonable towards their suppliers with respect to the statistics/data. There are no personal information or privacy issues related to the statistics/data and access by e.g. researchers is neither subject to any confidentiality or ethical constraints.

"I feel that understanding has increased, but then, we can get new suppliers and they may not be accustomed to having to provide statistics / data on our behalf. We need to look at whether it is ethical or not."

ICA

The ethical issues ICA faces are mostly related to how to track their consumers and how to supply and market their products. Consumer generated data can by law not be stored for more than 18 months and the use and level of detail of such data to create targeted discounts must be discussed. ICA experiences that their consumers generally don´t like

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

sharing data, but they do appreciate targeted discounts. Cigarettes and similar products are however not promoted nor subject to discount campaigns. Furthermore, when consumers use their loyalty ID, ICA has the possibility to look up the consumer, but this is only used to answer customer specific questions on e.g. bonuses. In addition, ICA analysts are not allowed to go into the ICA bank to get information on specific consumers.

"It is possible to do personal targeting or for some groups. Then again, we have to have a discussion about to what detail. How much do we want to do this? This is still a question. My colleague said we have some internal policies, e.g. we don´t give discounts on products like cigarettes. I don´t think we do it on beer either."

GfK

According to the GfK data acquisition is done in a transparent manner and ensures that all the selected household panellist give consent for participation. The organization guarantees the household panellist that their personal information provided will be protected in compliance with the Danish Law. Legal restrictions on personal data disclosure are strictly complied by the organization and blind any form of identifiable information contained in the dataset made available to businesses. The organization does not disclose the full details of household panellist to researchers because it is prohibited by the law. If there is need to provide household identifiable information such as the address, names, and other personal information, then a formal request has to be made to every household that participated in the survey for their consent and approval.

"We have always blinded the personal data they provide to us, phone numbers, and names, address. None of our customers, none of our consultants know any of the households. It is only the people managing households that know the household names. It is two people in Denmark. The manager has access to it as well to household information. So everybody else who works with data only know the household identification number and none of the detailed information will ever leave GfK."

SD

SD has obligation to protect any sensitive information provided by survey respondents. The turnover reported by the individual businesses is considered as sensitive information and it is protected in accordance to Danish Law. The informant believes that this information when provided could be used to identify the organization and it use for other purposes than which it was collected.

"Somehow the data are very sensitive they are confidential this is about turn over and how it is distributed we assume that by law we should take care of this data and ensure that individual do not misuse it."

Discussion

The discussion covered under this section is based on the key themes highlighting informants' perceptions related to the research questions. These themes are discussed in

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

line with the three main dimensions proposed to be of relevance in support of the development of Richfield research infrastructure as follows:

Relevance to scientific cases

Diversity of data sources. An important finding in terms of potential scientific reach is the diversity of data sources available for the proposed RI. As can be seen from the company descriptions, the RI may incorporate data from example retailers, public procurement companies, statistical institutions and market organizations. With such diversity, the proposed RI is more likely to successfully undertake e.g. broad EU-level socio-economic and consumption behaviour related research questions. With different kinds of data providers of different sizes, the RI ensures sufficient inter- and intra-regional data coverage with different levels of data resolution in the EU. The differences and reasons for data generation may also affect the scientific case. Looking at SD and Gothenburg City, data generation is significantly governed by policies to control suppliers or follow organic procurement which may put limits on the broadness and the potential usefulness of the data for scientific purposes. On the other hand, ICA and GfK may collect data from a broader consumer perspective, thus reflecting a more generic usage potential. It is thus important for the proposed RI to be aware of differences in data generation purpose and influence relevant data providers according to the needs of the scientific case. The RI needs to ensure data source diversity in collected data, mainly due to differences in reasons for data collection and associated generic data usage potential for EU inter- and intra-regional research.

Incompleteness of data. A key point with respect to the scientific case is also possible incompleteness of data from potential data providers. The proposed RI may need to encourage certain data management procedures from their data providers if relevant to the scientific case. For example, Gothenburg City has different levels of data aggregation depending on the specific contract, in order to avoid handling too big data files. If, instead, the raw data could be supplied to the RI, the relevance to the scientific case may be higher. Further, data is only available four years back in time as of 2016, which creates a time frame constraint in terms of a potential scientific case. The RI needs to ensure completeness of supplied data, preferably raw data, possibly by data management feedback to data suppliers.

Duration of data storage. A related problem at ICA is the allowed storage of business generated consumer data for eighteen months, according to Swedish law, which also adds to the limitations on the investigated time frame in the scientific case. The latter illustrates a potential cross-country legislation problem that has to be solved in relation to the RI. On the other hand, ICA may potentially provide both raw and aggregated data, according to the scientific needs. In the case of GfK, there may be a significant resolution problem in the captured data for the scientific case, since outside household purchases are not accounted for. Similarly, for SD, it may be important that the whole picture is considered, also including foodservice sales in the data. The time-frame for business data storage needs to be harmonized through a joint legislation procedure within the RI.

Missing or erroneous information in the provided data is something the RI must consider with respect to its use for scientific purposes. For example, ICA pointed to the fact that they

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

have a problem to trace errors due to their large assortment. On the other hand, for GfK and SD, the data is cross-checked or error corrected in order to avoid false data. RI users (e.g., scientists, researchers and policy makers) that are not directly linked to the data generation process are more likely to question its trustworthiness. This may be due to the potential uncertainty in data from different data providers. The best practice is to ensure that the data are corrected from the beginning and throughout collection processes. Further it is important that the RI has a rigorous scrutiny around provided data. The RI needs procedures for scrutinization of supplied data. As, the data may contain uncertainties, it is important that the RI has sufficient insight into the data collection processes.

Privacy issues. If the scientific case considers a detailed analysis involving tracking of single individuals, the proposed RI may have a legislation problem. As illustrated in the case of ICA, the organization cannot share their consumer ID loyalty data with e.g. the ICA bank in Sweden, and similar problems may therefore be expected to occur on the EU level. Thus an EU-legislation for intra-EU consumer identification may potentially be needed to solve such issues. Further, the allowed time frame to store consumer data with respect to consumer privacy, needs to be taken into account. Consumer privacy legislation concerning identification of consumers and storage of associated data needs to be considered in relation to the RI.

Relevance to user group strategies

Potential sharing benefits. This research revealed a variety of users for the generated data across the cases, for Gothenburg City and SD the stakeholders ranges from journalists, interest groups, authorities to research institutions. As potential data providers to the current platform it may thus be expected that the same stakeholders will be interested in the activities of the proposed RI. On the other hand, in contrast to ICA and GfK, stakeholders like Gothenburg City and SD seem more likely to freely share their data with the proposed RI. SD also expressed that their possible interaction with the RI can generate new insights on how their data can be used to address societal health issues and thus work more pro-actively to support the researchers' needs. Different data providers may be more likely to share their data freely with the RI.

Concerning ICA, it was pointed out that ICA appreciates research and since they are continuously looking at opportunities to increase the quality and output from their data analysis, such activities may be facilitated by a close collaboration with the proposed RI in joint consumer behaviour and health related research. In contrast, GfK explicitly perceives health issues to be the researchers' responsibility, sells data to research universities, but at the same time shares data for research purposes at the EU level. Thus, the proposed RI may need to conceive of different business models in these cases. The RI may need to consider different business models, e.g. direct payment for used data, with different data providers, where the mutual benefit aspect must be taken into account.

Potential sharing barriers. Business organizations, such as GfK will be more reluctant to share the generated data on RI than the public institutions knowing that retailers could potentially benefit from such data while not sharing their own data. As both a data provider and potential data user, it is interesting to look at how ICA and other retail organizations may interact with the proposed RI.

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

ICA is dependent on business generated consumer data in order to optimize their market performance, but they also buy complementary market data on competitors' performance, as well as selling it to suppliers and organizations. ICA is also willing to share data with universities for mutual benefit purposes. If the RI does not want to buy data from retailers, the mutual benefit aspect is something that the proposed RI needs to consider. Therefore, it is important that the RI can add another dimension to the provided data than the companies can already do themselves. This can be done by e.g. sharing more complete joint data sets in exchange for company-generated data. In addition, best practices will be to develop strategies that could help exclude entities, businesses organisation likely to have competitive urge over their counterparts sharing data on the RI to further the course of researchers' effort to address societal challenges. This can be considered in terms of paying for any data they use or allowing to use data that may not have competitive advantages. The RI may have difficulties in obtaining and handling complete data sets that contain sensitive information in terms of tracing companies' or competitors' performance.

Relevance to access strategies

Access permission. As can be concluded from the four interviews, that there are significant differences in how the RI can access the provided data. In the case of Gothenburg City, data access is only granted people within the organization, which in turn can provide the necessary data. Similarly, for GfK, the data is extracted from their platform, once an agreement on terms and conditions has been signed. For SD, data access requires affiliation with an authorized Danish institution which can assume the overall responsibility. Without an overall legislation around simplification of data sharing with EU-level research institutions, it may thus be expected that the RI may have to negotiate terms and conditions for data sharing with the different data providers in the different countries. The RI needs an overall EU legislation around data sharing with EU-level research institutions in order to avoid having to negotiate terms and conditions with different data providers in different countries.

Access agreement. A key issue with respect to the data sharing terms and conditions agreement, at least for SD, seems to be careful handling of non-aggregated data, in order to avoid possible identification of an associated enterprise. ICA also expressed this concern, since the company is exposed to the stock market and sensitive information on company performance can be used for non-legal money making purposes. Data access in specific projects thus has to be discussed within the organization. It is however more likely that ICA can share partial data with the RI, since this was considered less of an issue. A similar situation was also identified by GfK, where e.g. a retail business may acquire information from GfK that can be used to attract competitors' customers. SD also acknowledge problems with sensitive information e.g. business turnover, which is protected by Danish Law. Thus, the proposed RI is likely to require specific EU policies on data handling, data aggregation and disclosure to meet the concerns of data providers. The RI needs specific EU policies on data handling, data aggregation and disclosure to handle non-aggregated data which allows identification of companies and their associated customer base.

Access privacy. Regarding access privacy issues it is interesting to note differences between the interviewed cases. Gothenburg City sees no privacy issues related to the statistics/data

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

and access by e.g. researchers is neither subject to any confidentiality or ethical constraints. On the other hand, GfK guarantees that the personal information related to the household panellist will be protected by the Danish law. Thus, the organization may not be able to disclose the full details of the household panellist to the proposed RI, unless given active consent from each individual household. It seems like further EU legislation may be needed in this matter. The extent of the data privacy issue will likely depend upon specific data provider, provided data and country legislation and may be simplified through EU legislation.

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

Conclusions

The research was conducted with the aim to explore key stakeholders' opinions on what could be considered as best practice of generating data for extraction of consumer food purchasing and procurement behaviour. The themes identified shaped our discussions in the three main dimensions relating to the scientific case, user and access strategies. These are of great relevance for best practices that will improve our understanding on how to address potential challenges or barriers likely to encounter in developing RI. The table 2 Summarise the key themes and conclusions regarding the best practice identified for the RI development

Summary of key themes and conclusions

Diversity of data sources	The RI needs to ensure data source diversity in collected data, mainly due to differences in reasons for data collection and associated generic data usage potential for EU inter- and intra-regional research.
Incompleteness of data	The RI needs to ensure completeness of supplied data, preferably raw data, possibly by data management feedback to data suppliers.
Duration of data storage	The time-frame for business data storage needs to be harmonized through a joint legislation procedure within the RI.
Missing or erroneous information	The RI needs procedures for scrutinization of supplied data. As, the data may contain uncertainties, it is important that the RI has sufficient insight into the data collection processes.
Privacy issues	Consumer privacy legislation concerning identification of consumers and storage of associated data needs to be considered in relation to the RI.
Potential sharing benefits	RI needs to demonstrate the value creation potential which could be the benefit to data providers/users.
Potential sharing barriers	Data providers' potential revenue loss and competitive disadvantage will hinders sharing and should be addressed by RI business model.
Access permission	The RI needs an overall EU legislation around data sharing with EU-level research institutions in order to avoid having to negotiate terms and conditions with different data providers in different countries.

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

Access agreement	The RI needs specific EU policies on data handling, data aggregation and disclosure to handle non-aggregated data which allows identification of companies and their associated customer base.
Access privacy	The extent of the data privacy issue will likely depend upon specific data provider, provided data and country legislation and may be simplified through EU legislation.

The generated data is an important component of the RI, without it the development and the existence of the RI will be valueless and therefore the need to identify the appropriate types of data within the research domain of food, consumption and health. The RI will be fed with huge diverse data which can be harnessed by potential users, researchers for broader EU level socio-economic and consumption. The vastness of data can either be potential and hindrance making it necessary for stakeholders especially the scientific community to define data types needed for RI. As data remains a vital resource for RI, our findings indicate that current data coverage in some cases are limited. Different data needs were identified by the case organisations in question and the essence to capture it in order to broaden its use for research was emphasised. It is important that RI should have measures in place to help identify providers' data needs deemed important for the scientific case. This could be useful information for potential users of the platform to search for complimentary data or enter into dialogue with providers to capture such data.

Key message 1: RI should establish minimum data holding duration and be able to negotiate for exemptions on data expiration or holding extension purposely for research.

Depending on the type of data collected, different regulations dictate the extent of duration it could be stored for use. This implies that data could only be available to RI for a defined period of time. Hence RI user's data demands extending beyond this period could be comprised. It would be necessary to identify data that falls within this category and negotiate for exemptions on data expiration or holding extension purposely for research.

Key message 2: RI data requirements should lay emphasis on measures that address data provider's error correction strategy.

As RI will depend on data generated from diverse sources and it could be the basis for end users such as researchers to question the validity and trustworthiness of the data. RI data requirements should lay emphasis on measures that address data provider's error correction strategy. For instance RI should require data providers to clearly define their dataset with all needed information such as metadata to facilitate user trustworthy.

Key message 3: Data sharing collaboration should permit data provider to access their own data and other database in RI and also get feedback on usage of their dataset.

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

The willingness of the case organisations to share data for RI is highly related to the possible benefits they stand to gain. The mutual benefit of sharing data to address societal challenges seems to be the common grounds for data sharing collaboration. This could be considered in terms of permitting data provider to access their own data and other database in RI and also get feedback on usage of their dataset. For the business organisation like GfK this might not be enough because of the fear of losing revenue as a result of competitors or clients using their data free of charge.

Key message 4: RI business model should help data providers gain financially from organisations or institutions using their data for commercial purpses.

As a matter of urgency, RICHFIELDS must strategize and should help these organisations gain financially from organisations or institutions using data for commercial purposes. On the other hand, data sharing could be possibly hindered by access restriction policies, such as whom and what the data can be used. Appropriate access arrangements and agreed standards with the data providers should be instituted to reduce data access restriction and facilitate potential users' collaborative access right to the RI.

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

References

1. Brammer, S. and Walker, H. (2011) Sustainable procurement in the public sector: an international comparative study. International Journal of Operations & Production Management, 31(4), pp. 452-476.
2. Lehtinen, U. (2012) Sustainability and local food procurement: a case study of Finnish public catering. British Food Journal, 114 (8), pp.1053 – 1071
3. Johnston, J.L. et al. (2014) Understanding Sustainable Diets: a descriptive analysis of the determinants and processes that influence diets and their impact on health, food security, and environmental sustainability. Advances in Nutrition 5, pp. 418-429.
4. Mikkelsen, B.E. (2012) Organic Foods on The Public Plate – Technical Challenge or Organizational Change? J Foodservice Business Research 15(1).
5. McCrudden, C. (2004). Natural Resources Forum 28 257–267 Using public procurement to achieve social outcomes
6. UNEP (2012) 'Sustainable Public Procurement Implementation Guide: Introducing UNEP's approach.' United Nations Environment Programme, Paris: France.
7. Stefani, G., Tiberti, M. and Lombardi, G.V. (2015) Public Food Procurement: A Systematic Literature Review Working Paper N. 2/2015
8. Sonnino, R and Morgan, K. (2007) The European School food revolution.
9. Rimmington, M. et al. (2006) Corporate social responsibility and sustainable food procurement. British Food Journal, 108 (10), pp. 824-837.
10. Mørk, T., Tsalis, G. and Grunert, K.G. (2014) Økologi I Offentlige Køkkener (Organics in public kitchens).
11. Barling, D. et al. (2013) 'Revaluing Public Sector Food Procurement in Europe: An Action Plan for Sustainability', available at: http://www.foodlinkscommunity.net/fileadmin/documents_organicresearch/foodlinks/publications/Foodlinks_report_low.pdf
12. Romani, A. Q., Lund, R. L and Mikkelsen, B. E. (2016). Does removal of the VAT on fruit and vegetables increase sale–results from a price manipulation intervention on food sales in a Danish retailer. Journal of Behavioral Nutrition and Physical Activity.
13. Pawlowski, K.D and Lindum, J. (2016) The Effect On Sound Nudging on Customer Behaviour and Vegetable Sales in a Supermarket, Abstract for ICCAS 2017
14. Ørnbo, L. E., Tvedebrink, T. D. O., Sudzina, F. & Mikkelsen, B. E Slightly reducing salience of sugar sweetened beverages reduce sale in vocational school canteens, Abstract for ICCAS 2017
15. Yin, R. K. (2003). Case study research design and methods third edition. Applied social research methods series, 5.
16. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), 77-101.

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

Appendix 1: RICHFIELDS interview guide for WP 8 interviews

Introduction

Welcome and thanks for agreeing to be an interviewee.

Purpose of this interview

We are conducting this interview as part of the Richfield Project. The WP 8 of that project aims at investigating best practices cases of extracting purchasing and procurement behaviour from existing data sources. It also studies how new technologies and devices of the farm to fork supply chain can be used to source data from the retail and food service sector and how they can be utilized by the RI Consumer Data Platform. It finally aims at discussing how such data might be used in future research infrastructures in the borderline between food and health science.

The purpose of this interview is to elicit stakeholders view on

(1) what constitutes best practices for businesses to collect food purchasing/procurement data,
(2) how can these data be used
(3) how these data are structured.
(4) how IC technologies can be used to capture food purchasing/procurement data.
(5) if such data can be of value to research community
(6) in that case what the privacy issues/IPR and/or ethical constraints might be

Interview questions

Section A: Best practices of capturing food purchasing/procurement data

1. For what purpose does your organization collect food purchasing/ procurement data?

2. How are these data structured?
    a. Would you explain why the data is structured in this format?

3. What do you consider to be the content of the data captured?
    a. Could you elaborate on the relevance of the data content[i]?

4. Can you explain how the data capture is used to generate insights to benefit the "retail industry"?
    a. What would you consider to be the significant challenge(s) of using the food purchasing/procurement data to the benefit of the industry

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

5. What information needs are not captured in the current way of data collection?
    a. How do you think this challenge can be addressed?

6. What is your opinion about using the food purchasing/procurement data to address societal challenge, “eg. Prevention of diet-related diseases”?

7. Can you think of cases where your data could be shared with the research community eg Richfield platform?

    a. What benefits do you think it can bring to the research community providing such data?
    b. What benefits do you think it can bring to other stakeholders providing the data
    c. What should be done to promote sharing of data for research purposes?

8. What privacy policy covers the capturing and sharing of the purchase/procurement data?
    a. Can you elaborate on any legal restrictions on data sharing for research purposes?

9. Could you suggest topics (ethics) to be discussed in Richfield project?
    a. Why do you consider it to be important

[i] A note on difference between structure and content: Structure is more superficial and relates to how the data are put together or displayed. For example a data can be structured in text form, tabular or even graphical form or format but the content is more related to its meaning.

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