Manuscript Details

Manuscript number	GFS_2019_29
Title	An overview of UK household datasets to monitor transitions to sustainable diets
Article type	Invited research article

Abstract

There is growing international consensus that current patterns of food consumption are not sustainable and global change is needed. Understanding the mechanisms for a transition towards more sustainable diets requires systematic monitoring at the individual or household level, together with temporal sampling. Many countries collect panel data on food expenditure and consumption, but it is not clear if they are suitable to develop a clear understanding of how or why diets are transitioning to become more or less sustainable. Our aim is to identify and describe existing food and diet datasets available in the UK and to assess the extent to which they can be employed to monitor transitions to sustainable diets. We show the UK has a large number of datasets tracking individual or household food purchases and consumption over time. However, current data sources are not suited to gain insight into how and why individuals are (or are not) transitioning to sustainable diets. With the exception of proprietary datasets, most datasets only collect data annually, making it challenging to understand fine-scale behavioural change over shorter timeframes. Thus, there is an opportunity to design and implement an open access UK sustainable diets data collection effort at the household level. These efforts can be complemented with recent innovations in data science methods and digital technologies – such as dietary intake trackers – that along with supporting individuals in their dietary behaviour change may enable collection of high quality datasets.

Keywords	Panel data; food consumption; sustainable diets; data science; digital technologies; review.

Corresponding Author	Christian Reynolds
Corresponding Author's Institution	The University of Sheffield
Order of Authors	Remco Benthem de Grave, Niki Rust, Christian Reynolds, Anthony Watson, Jan D Smeddinck, Diogo Souza Monteiro
Suggested reviewers	Rosemary Green, Peter Horton

Submission Files Included in this PDF

File Name [File Type]

Highlights.docx [Highlights]

Front page with highlights.docx [Title Page (with Author Details)]

draft manuscript gfs-clean_3006 BLIND.docx [Manuscript (without Author Details)]

Table 2 for paper.docx [Table]

Conflict of interest.docx [Conflict of Interest]

To view all the submission files, including those not included in the PDF, click on the manuscript title on your EVISE Homepage, then click 'Download zip file'.

Highlights:

- Individually each UK dataset has limited effectiveness to monitor transitions to sustainable diets.
- No single dataset recorded purchased and consumed quantities, along with attitudes/perceptions of sustainability and food consumption or purchase.
- Multiple UK datasets can be used to collectively conduct analyses of general trends and to compare different cohorts regarding the changes toward sustainable dietary patterns.
- Not all UK datasets are linked to databases containing environmental impact information. Though this linkage is currently occurring.
- New technology can improve assessment of changes towards sustainable diets. Including digital wearable devices to collect data on food choices, and novel data science methods.

Proposed title: An overview of UK household datasets to monitor transitions to sustainable diets

Remco Benthem de Grave, Institute of Neuroscience, Newcastle University, UK r.benthemdegrave2@newcastle.ac.uk

Niki A. Rust, Newcastle University, UK niki.rust@ncl.ac.uk

Christian J Reynolds, Department of Geography, University of Sheffield, and the Barbara Hardy Institute, University of South Australia c.reynolds@sheffield.ac.uk, ORCID 0000-0002-1073-7394 (Corresponding author).

Anthony W. Watson, Human Nutrition Research centre, School of Biomedical Sciences, Newcastle University UK. anthony.watson@newcastle.ac.uk

Jan D Smeddinck, Open Lab, Newcastle University jan.smeddinck@ncl.ac.uk

Diogo M. de Souza Monteiro, Centre for Rural Economy, School of Natural and Environmental Sciences, Newcastle University, diogo.souza-monteiro@ncl.ac.uk

Abstract: There is growing international consensus that current patterns of food consumption are not sustainable and global change is needed. Understanding the mechanisms for a transition towards more sustainable diets requires systematic monitoring at the individual or household level, together with temporal sampling. Many countries collect panel data on food expenditure and consumption, but it is not clear if they are suitable to develop a clear understanding of how or why diets are transitioning to become more or less sustainable. Our aim is to identify and describe existing food and diet datasets available in the UK and to assess the extent to which they can be employed to monitor transitions to sustainable diets. We show the UK has a large number of datasets tracking individual or household food purchases and consumption over time. However, current data sources are not suited to gain insight into how and why individuals are (or are not) transitioning to sustainable diets. With the exception of proprietary datasets, most datasets only collect data annually, making it challenging to understand fine-scale behavioural change over shorter timeframes. Thus, there is an opportunity to design and implement an open access UK sustainable diets data collection effort at the household level. These efforts can be complemented with recent innovations in data science methods and digital technologies – such as dietary intake trackers – that along with supporting individuals in their dietary behaviour change may enable collection of high quality datasets.

Keywords: Panel data; food consumption; sustainable diets; data science; digital technologies; review.

Highlights:

- Individually each UK dataset has limited effectiveness to monitor transitions to sustainable diets.
- No single dataset recorded purchased and consumed quantities, along with attitudes/perceptions of sustainability and food consumption or purchase.
- Multiple UK datasets can be used to collectively conduct analyses of general trends and to compare different cohorts regarding the changes toward sustainable dietary patterns.
- Not all UK datasets are linked to databases containing environmental impact information. Though this linkage is currently occurring.
- New technology can improve assessment of changes towards sustainable diets. Including digital wearable devices to collect data on food choices, and novel data science methods.

An overview of UK household datasets to monitor transitions to sustainable diets

Abstract: There is growing international consensus that current patterns of food consumption are not sustainable and global change is needed. Understanding the mechanisms for a transition towards more sustainable diets requires systematic monitoring at the individual or household level, together with temporal sampling. Many countries collect panel data on food expenditure and consumption, but it is not clear if they are suitable to develop a clear understanding of how or why diets are transitioning to become more or less sustainable. Our aim is to identify and describe existing food and diet datasets available in the UK and to assess the extent to which they can be employed to monitor transitions to sustainable diets. We show the UK has a large number of datasets tracking individual or household food purchases and consumption over time. However, current data sources are not suited to gain insight into how and why individuals are (or are not) transitioning to sustainable diets. With the exception of proprietary datasets, most datasets only collect data annually, making it challenging to understand fine-scale behavioural change over shorter timeframes. Thus, there is an opportunity to design and implement an open access UK sustainable diets data collection effort at the household level. These efforts can be complemented with recent innovations in data science methods and digital technologies – such as dietary intake trackers – that along with supporting individuals in their dietary behaviour change may enable collection of high quality datasets.

Keywords: Panel data; food consumption; sustainable diets; data science; digital technologies; review.

Highlights:

- Individually each UK dataset has limited effectiveness to monitor transitions to sustainable diets.
- No single dataset recorded purchased and consumed quantities, along with attitudes/perceptions of sustainability and food consumption or purchase.
- Multiple UK datasets can be used to collectively conduct analyses of general trends and to compare different cohorts regarding the changes toward sustainable dietary patterns.
- Not all UK datasets are linked to databases containing environmental impact information. Though this linkage is currently occurring.

1. Introduction

Current food purchase and consumption patterns are leading to unhealthy diets (Kearney, 2010), which in turn are linked to increased prevalence of non-communicable diseases, such as obesity, type 2 diabetes and cardiovascular disease (Aston et al., 2012; Blundell and Cooling, 2000). Moreover, there is mounting evidence that the production, processing, transport and final preparation of food to support current dietary patterns has increasing environmental costs and is unsustainable as it leads to increasing eutrophication, greenhouse gas emissions, as well as land and biodiversity loss (Aleksandrowicz et al., 2016; Green et al., 2015; Poore and Nemecek, 2018; Tilman and Clark, 2014; Willett et al., 2019).

As the evidence of the contribution of food production and consumption to the deterioration of planetary health becomes clear, so does the need to help consumers choose more sustainable diets (Willet et al, 2019). The FAO (Burlingame and Dernini, 2012) and the first and second US National Academies of Sciences, Engineering and Medicine Workshops on Sustainable Diets, Food and Nutrition (Institute of Medicine, 2014; National Academies of Sciences, Engineering, and Medicine et al., 2019) suggest that sustainable diets should be affordable and acceptable, healthy and nutritionally balanced and with low environmental impact. Transitioning towards sustainable diets is directly related to all of the United Nations Sustainable Development Goals[1]. In summary: there is a clear need for rapid, international change in how we produce and consume food. Importantly, changes in demand patterns, will eventually lead to changes in production (Horton, 2017). In fact, Ingram (2017) argues that we need to change the way we look at food systems and, rather than emphasizing the need to increase production, we should focus on managing demand.

A common concern considering transitions to more sustainable diets is how to encourage rather than impose or force change. Research in this area faces a diverse landscape of policies and interventions that are tailored to specific individuals or cohorts. It is therefore important to systematically monitor how effective different interventions have been and how transitions are occurring. To these ends, data needs to be gathered at the individual or household level with a regular frequency to observe what is being purchased, prepared, and consumed both at home and away from home.

In many developed countries there are both private and public data collection efforts collecting and recording information on food expenses, consumption patterns and nutrition[2]. De Keyzer et al. (2015), Perignon et al. (2017), as well as Bandy et al., (2019) have conducted systematic reviews of food consumption datasets and found important gaps and limitations regarding the applicability of these datasets for monitoring transitions to sustainable food consumption behaviour. Highlighting differing definitions of "sustainability" (such as being affordable, acceptable, healthy and/or low in environmental impact), the reviews found that at most two dimensions of sustainability are captured in the studies based on the datasets they reviewed, and that the majority of the datasets contains only one such dimension. These reviews indicate that there does not appear to be a systematic data collection effort capturing all the dimensions of "sustainability". However, databases currently exist that allow estimating nutritional values, greenhouse gas emission (GHGE) and cost

[1] https://www.stockholmresilience.org/research/research-news/2016-06-14-how-food-connects-all-the-sdgs.html

[2] For example the World Bank Global Consumption Database compiles food expenditures across food and drinks expenses from a nationally representative sample of developing countries households (.http://datatopics.worldbank.org/consumption/sector/Food-and-Beverages). Similar datasets are available from international organizations like the OECD, the European Union and the national statistics of all high income countries.

from purchased or consumed products (see for instance www.ggdot.org, (Hobbs et al., 2015; Horgan et al., 2016; Monsivais et al., 2013) etc). Thus, there may be opportunities to use such databases in along with existing comprehensive datasets on food purchasing and consumption patterns to understand the evolution of sustainable diets without the need of new data collection efforts.

We provide an overview of existing private and public datasets of food purchases and consumption patterns in the UK and discuss their suitability to assess transitions and changes towards sustainable diets. The UK is an interesting starting point and case for observation with respect to this topic because the sustainability of current diets has been questioned (Reynolds et al., 2019; Reynolds et al., 2015; Wrieden et al., 2017). The UK is committed to meeting the UN Sustainable Development Goals, recently also declaring the goal of reaching 'net zero' carbon emissions by 2050 (Pye et al., 2017; Walker et al., 2019), and has an actively engaged political and civil society in developing approaches to improving the current dietary and environmental situation. Moreover, the UK has a strong tradition and capacity to collect data on food purchase and consumption (Oddy, 2003; Orr, 1937).

As an addition to providing a comprehensive overview and discussion of available datasets on UK food purchases and consumption patterns, to support future data collection efforts, we also provide suggestions for approaches to improving the completeness, quality, and linking of existing datasets, as well as the potential for improved data collection and monitoring with digital technologies. As such, next to informing further research, this work provides guidance and evidence on improving data collection that can lead to improved monitoring and understanding of transitions towards more sustainable diets. The outcomes can therefore be helpful to policy makers, research an industry alike.

2. Methods

To identify existing datasets available in the UK, we followed a similar search strategy to that employed by. Blanquer et al. (2009) and De Keyzer et al. (2015). Primarily, we searched for datasets that allowed researchers to reconstruct a complete diet of individuals or households with multiple time spaced diet assessments. This was the basis criterion as a range of sustainability dimensions can be estimated if this diet information is available; healthiness of diets can be estimated using nutrition profile tables, affordability can be partly assessed using food cost tables, GHGE can be assessed through conversion tables, and acceptability, safety and accessibility of diets can be reasonably assumed as these are self-selected diets.

To be able to reconstruct a complete diet, we defined that the data should cover at least one complete consumption day (e.g. through a 24h Dietary Recall (24h-DR), a Diet Diary (DD), an extensive Food Frequency Questionnaire (FFQ), or a purchase diary of at least a week).

We further focused on data that was collected in the UK and is available for research purposes. We excluded datasets that focus exclusively on children or the very old, as well as datasets that consist of secondary data collection efforts (i.e. merging data collection efforts done elsewhere).

A first list of datasets was created from authors combined knowledge of (publications about) data collection efforts describing diets in the UK. Next we consulted the UK data service (see https://www.ukdataservice.ac.uk/get-data/themes/food.aspx). To identify any additional datasets, we contacted researchers through personal networks who are doing empirical analysis of food consumption. We also reached out to private companies that collect diet information (but not necessarily in the UK) and to experts groups such as the Food and Climate Research Network (FCRN) google group (https://groups.google.com/forum/#!topic/fcrn-l/TRMs4BnUWYc).

For each dataset identified, we collected characterising information and metadata from the description that comes with the dataset. In some cases, we had to refer to the original survey questionnaires, to the raw data, or to publications that use the dataset. Institutions were contacted to verify entries and asked for missing information, though some data holders did not return answers.

3. Results

This section presents the datasets we were able to identify. In table 1, we list the majority of datasets that were investigated for inclusion in our overview. In table 2, we describe, in detail, the 9 datasets that meet our inclusion criteria. All of these datasets fulfil the aforementioned requirement of providing a complete overview of at least one day of consumption or purchase data and can be accessed for research purposes (although some need to be purchased).

Table 1: List of British household panel datasets gathering data on food expenditure and consumption

Dataset or survey name	Public		Private	
	Open	Restricted	Fee	Restricted
EPIC Norfolk (Day et al., 1999)		√		
EPIC Oxford (Davey et al., 2003)		√		
Family Food module of Living Cost and Food Survey (LCFS) (Department For Environment and Office For National Statistics, 2017) (Office For National Statistics, 2019)		√		
Fenland study ("Fenland Technical Summary - MRC Epidemiology Unit," n.d.)		√		
Kantar consumption panel			√	
Kantar purchase panel			√	
National Diet and Nutrition Survey (NDNS) (Laboratory and Research, 2019)		√		
UK Women Cohort Survey (UKWCS) (Cade et al., 2015)		√		
UKBiobank (Sudlow et al., 2015)	√			
Health Survey for England		√		
1000 family study		√		
85+ study		√		
ASH30		√		
ALSPAC		√		
FAO statistics	√	√ [2]		
Food and Drink in Scotland		√		
Gateshead Millennium Cohort		√		
GfK (company)			?[1]	
Global Dietary Database (GGD)		√		
Loyalty card data collections (e.g. Dunnhumby, Tesco, Sainsbury, Waitrose)			√	√[3]
MyFitnessPall (company)				√
Nielson (company)			?[1]	
Scottish Health Survey		√		
Slimming world (company)				√
Weightwatchers (company)				√

[1] Data for the UK for these companies may not be available, but this was not conclusively verified (the companies did not respond to an information request). [2] Greater detail available via application for restricted data for some areas.[3] Some Loyalty card data available through UKDS and the CDRC.

In table 1 in bold are the datasets for which we have meta-data and are further described in table 2. The datasets have been categorized into public datasets (those collected by governmental agencies or funded by public research funds) or private sources (those collected by commercial companies, generally through apps, surveys, home or retail scanners). Public datasets are divided into open or restricted, meaning that further access permissions where institutional associations need to be verified and sometimes special permission request need to be provided. Private datasets are divided

into those that are available for a fee and those that are generally not shared outside the company (restricted private datasets).

Table 2 below presents and characterizes the 9 datasets that met our main criteria. Next we briefly explain the characteristics of these data in three dimensions: sampling and recruitment, data collection methods and economic information therein.

<<<<Table 2 here>>>

Study design, recruitment and sample characteristics

Three types of designs can be recognized in the overview. First, two of the nine studies concerned non-cohort studies (National Diet and Nutrition Survey, and the Family Food Module of Living Cost and Food Survey, FFM-LCFS). Both of these cross-sectional studies targeted UK households using a multistage stratified sampling strategy in which households were identified from Postcode Address Files (PAF) and recognized as small users, and clustered in Primary Sampling Units (PSUs). Households were then drawn from a number of PSUs. Samples sizes ranged from about 1000 participants annually in the NDNS to 6000 households annually.

Second, five datasets concerned cohort studies with a clear time limited design (EPIC Norfolk, EPIC Oxford, the Fenland Study, the UKBiobank and the UK Women's Cohort Survey, UKWCS). Targeted populations varied considerably. Some studies targeted specific diets (non-red-meat-eating, vegetarian), some geographical regions (Norfolk, Cambridgeshire) and one study targeted women only. All studies targeted a middle age range however, with participant ages ranging from 20 to 79. NHS registers and membership lists (e.g. that of the vegan society) where used to recruit people. Cohort sizes of ranged from roughly 12,500 (Fenland Study) up to roughly 211,000 (UKBiobank), although sample sizes at the level of individual recordings range from 1600 to 100,000.

Third, the data collected by Kantar. These are the only commercial datasets and the only datasets that monitor participants' diets over an unrestricted time frame (4x per year with 10,000 people in the consumption panel and 30,000 people in the purchase panel). Advertisements on social media were used to recruit people, although more targeted methods were also used to obtain a representative sample size.

Dietary assessment methods, administration method and method of portion size estimation

A variety of methods to assess dietary consumption or purchases can be found between and within the databases. These include Food Frequency Questionnaires (FFQ), 24-Hour Dietary Recalls (24h-DR), Diet Diaries (DDs), and purchase diaries.

Food Frequency Questionnaires (FFQs) were used in four studies. These questionnaires asked about habitual consumption frequency in the past 12 months on a range of food items (28 to 217 food items). Participants were requested to rate their consumption frequency from never, to 6 per day on 9 frequency choices. Some exceptions to this are that one study (UKWCS) used a 10-point frequency scale and two smaller FFQs in EPIC Oxford used a 6-frequency scale. Portion sizes were generally estimated by framing the question such that it asked for the consumption of standard portion sizes. The standard portion size was then described with the item or category. Some questionnaires omitted portion size and only asked for a frequency. We note that some of the

smaller FFQs do not describe a full consumption day. However, other assessments in the same study do. The small FFQ's were included for completeness.

The 24h-DR was used in three studies. These asked about the consumption of the previous day. Methods used varied from pen and paper recordings, accompanied with suggestions on standard portion sizes, to online forms that required to rate their portion sizes in standard measures. The 24h-DRs were all self-administered, either at the test centre or at home.

Diet Diaries (DDs) were used in five studies. These asked the participants to track their consumption for several days (ranging between studies from 4 to 7 days). In both EPIC studies and the NDNS paper, DDs were used in combination with suggestions for standard portion sizes, supported by pictures of various portion sizes that participants could refer to. In the UKWCS participants were asked to list weight or volume of consumed products which had to be measured or read from packaging (standard measures were allowed on some occasions). The DD in Kantar was performed on a computer. Participants selected per meal the products that they had used, but did not specify consumed amounts.

Purchase diaries where used in two studies. The FFM-LCFS used pen and paper entries or allowed participant to attach their receipts. In the Kantar purchase panel, participants were asked to scan each purchase receipt using a digital clicker. Both purchase diaries are self-administered and completed at home.

Economic information

Income is recorded for five out of the nine studies we describe (the NDNS, the FFM-LCFS, the Fenland Study and both Kantar datasets), while prices and/or expenditure are also recorded in the purchase panels (FFM-LCFS and Kantar datasets). Together, this allows the assessment of affordability.

4. Discussion

Above we have identified and described nine comprehensive datasets of diet, consumption, or food purchases in the UK that are available to the research community. Individually, each dataset has limited effectiveness to monitor transitions to sustainable diets and for direct comparisons between datasets. This is because they were not designed for either of these purposes. The datasets use different units of observation, sampling sizes[3], sampling rates, and study durations. In addition, none of the datasets recorded both purchased and consumed quantities (thus not allowing to estimate food waste (Reynolds et al., 2019)). In this regard, our outcomes are consistent with those of Perignon et al. (2017), who found that there is a lack of relevant and good-quality datasets for assessing the environmental, health and socio-economics impact of current diets.

However, we propose that collectively these datasets have the potential to assess transitions and changes towards sustainable diets in the UK. For these purposes, the identified datasets have to be linked to databases containing environmental impact information of the foods consumed or purchased. This is not currently the case for all identified datasets. This is linkage is possible, however, with mapping activities currently being undertaken for multiple datasets reviewed. This is labour-intensive to varying degrees, due to the different levels of food classification and dimensions for data-aggregation in each database. In order to scale the approach, methods for the automated the mapping and linking of dietary and environmental impact databases are required (Eftimov et al., 2017). At the same time, even if they are not linked directly to environmental

[3] For some datasets it is uncertain whether they present a representative sample of the British population.

impacts, these databases can still be used to collectively conduct analyses of general trends and to compare different cohorts regarding the changes in dietary patterns.

The datasets that were collected openly and which are publicly accessible have a snapshot nature and are suitable to evaluate how different groups have changed diets and facilitate cross sectional analysis. The value of the household food purchases panel data (such as Kantar) is that it enables researchers to observe transitions with a much finer granularity. We can compare how different households are changing consumption of a given food category and we can compare across food categories, household types, and time periods (weekly in the case of Kantar, or yearly e.g. for LCFS). In isolation, these datasets do not necessarily gather information on the health status of the household they recruit. However, since they contain detailed information on each product (or category) purchased that can be mapped onto datasets with information of the environmental impact of different households, and linked with income and prices to understand the tensions between sustainable and affordable diets. In addition, there is a lack of detail in current panel data on the traceability and origin of food; this additional information is needed to truly understand sustainability of different foodstuffs.

It should be highlighted that there is a certain degree of self-selection bias on the households that are included in both public and private panels that were reviewed. Moreover, these datasets have not inquired about households' attitudes to - or perceptions of - sustainable dimensions of food consumption or purchase (this would be required to understand reasons why people make changes in what they eat). Moreover, there is limited information about the context and practices with which the households live. Indeed, the food availability landscape is not necessarily captured in the datasets we have identified. However, those factors are important determinants of consumption and purchase. Consequently, there must be caution not to infer causality when interpreting data, for what may be the causes of any change in consumption/purchase patterns.

Still, the complementarity between the more frequent and rich information on products gathered in panel data and the broad coverage of large cohort studies presents a clear opportunity for assess general transitions to sustainable diets. The household panel data could be employed to identify trends and micro-responses to interventions, in turn the cohort studies can be used to confirm how they are impacting broader aggregate measures. Another opportunity lies with linking both private and public datasets to geographical information (which is recorded in differing detail in each dataset) to further our understanding of how changes in regional or urban food policies may be affecting consumption patterns, as well as environmental and health outcomes.

To overcome the aforementioned limitations of current datasets and to develop new datasets, we suggest harnessing technological developments to better assess dietary transitions and changes towards sustainable diets. We therefore briefly highlight the potential of digital wearable devices to collect data on food choices, as well as the use of data science methods to provide new methods of data harmonization and mapping.

In principle, data science methods (including frequentist statistics, probabilistic methods, as well as different techniques from machine learning and artificial intelligence) can be used for two main purposes with respect to the existing datasets: 1) improving the data-quality and reducing sparsity (filling gaps, e.g. data imputation (Jerez et al., 2010)), 2) linking datasets (e.g. through auto-correlation) ("Automated census record linking: a machine learning approach," n.d.), 3) clustering datasets or supersets, creating new sectioning or subsets (e.g. using autoencoders (Baldi, 2012)), 4) optimizing future / ongoing data collection (Sra et al., 2011) and 5) prediction.

At the same time, with the growing capabilities and affordability of sensors and increased computational capacity at hand and in the cloud, digital technology, including devices and software applications opens interesting opportunities for improving data collection and research efforts. Digital data streams can be very complex and have a high sampling rate – which can at times even emulate real-time "natural fidelity" recording, compared to what is feasible with more traditional data collection efforts. This area can be split into four main elements: 1) quantified self and community applications with a) self-reporting tools, such as consumption / intake trackers (Bradley et al., 2016), or b) habit tracking / forming apps (Stawarz et al., 2015), 2) general dietary information tools (Boulos et al., 2015), 3) professional practice support (Simons et al., 2012) and 4) indirect information sources (such as product sales data, raw materials uptake / tracking, supply-chain monitoring, distributed ledgers, as well as production and transport cost /energy expenditure monitoring).

5. Conclusions and future work

We conducted a review of existing data sources that can inform research on monitoring transitions towards more sustainable diets in the UK. As a key outcome, we presented an overview table of the available datasets and discussed the applicability for said analyses. We conclude that neither of the datasets fulfils the requirements for reliable monitoring or prediction. Most of the datasets are also limited to traditional data sources, such as survey responses. This clearly suggests two pathways for future work: improving the quality and linking the existing data sets, as well as a broader effort to collect coherent data on transitions towards more sustainable diets that combines systemic as well as individual-level data, including motivations and objective behaviour and consumption tracking. In both cases, digital technologies can play a key role and enable approaches that would not have been possible without them. This includes both software with supportive algorithms and user interfaces, which can, for example, gauge shopping behaviour, shopping, and the engagement with – and social communication about – diet information sources, as well as (sensing) hardware devices that allow for objective measurements e.g. of eating behaviour.

6. Bibliography

Aleksandrowicz, L., Green, R., Joy, E.J.M., Smith, P., Haines, A., 2016. The impacts of dietary change on greenhouse gas emissions, land use, water use, and health: A systematic review. PLoS One 11, e0165797. doi:10.1371/journal.pone.0165797

Aston, L.M., Smith, J.N., Powles, J.W., 2012. Impact of a reduced red and processed meat dietary pattern on disease risks and greenhouse gas emissions in the UK: a modelling study. BMJ Open 2. doi:10.1136/bmjopen-2012-001072

Automated census record linking: a machine learning approach [WWW Document], n.d. URL https://open.bu.edu/handle/2144/27526 (accessed 6.30.19).

Baldi, P., 2012. Autoencoders, unsupervised learning, and deep architectures. Proceedings of ICML workshop on unsupervised and transfer learning 37.

Bandy, L., Adhikari, V., Jebb, S., Rayner, M., 2019. The use of commercial food purchase data for public health nutrition research: A systematic review. PLoS One 14, e0210192. doi:10.1371/journal.pone.0210192

Blanquer, M., García-Alvarez, A., Ribas-Barba, L., Wijnhoven, T.M.A., Tabacchi, G., Gurinovic, M., Serra-Majem, L., 2009. How to find information on national food and nutrient consumption surveys across Europe: systematic literature review and questionnaires to selected country experts are both good strategies. Br. J. Nutr. 101 Suppl 2, S37–50. doi:10.1017/S0007114509990572

Blundell, J.E., Cooling, J., 2000. Routes to obesity: phenotypes, food choices and activity. Br. J. Nutr. 83 Suppl 1, S33–8.

Boulos, M., Yassine, A., Shirmohammadi, S., Namahoot, C., Brückner, M., 2015. Towards an "internet of food": food ontologies for the internet of things. Future Internet 7, 372–392. doi:10.3390/fi7040372

Bradley, J., Simpson, E., Poliakov, I., Matthews, J.N.S., Olivier, P., Adamson, A.J., Foster, E., 2016. Comparison of INTAKE24 (an Online 24-h Dietary Recall Tool) with Interviewer-Led 24-h Recall in 11-24 Year-Old. Nutrients 8. doi:10.3390/nu8060358

Burlingame, B., Dernini, S., 2012. SUSTAINABLE DIETS AND BIODIVERSITY DIRECTIONS AND SOLUTIONS FOR POLICY, RESEARCH AND ACTION.

Cade, J.E., Burley, V.J., Alwan, N.A., Hutchinson, J., Hancock, N., Morris, M.A., Threapleton, D.E., Greenwood, D.C., 2015. Cohort profile: the UK women's cohort study (UKWCS). Int. J. Epidemiol. doi:10.1093/ije/dyv173

Davey, G.K., Spencer, E.A., Appleby, P.N., Allen, N.E., Knox, K.H., Key, T.J., 2003. EPIC-Oxford: lifestyle characteristics and nutrient intakes in a cohort of 33 883 meat-eaters and 31 546 non meat-eaters in the UK. Public Health Nutr. 6, 259–269. doi:10.1079/PHN2002430

Day, N., Oakes, S., Luben, R., Khaw, K.T., Bingham, S., Welch, A., Wareham, N., 1999. EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. Br. J. Cancer 80 Suppl 1, 95–103.

De Keyzer, W., Bracke, T., McNaughton, S.A., Parnell, W., Moshfegh, A.J., Pereira, R.A., Lee, H.-S., van't Veer, P., De Henauw, S., Huybrechts, I., 2015. Cross-continental comparison of national food consumption survey methods--a narrative review. Nutrients 7, 3587–3620. doi:10.3390/nu7053587

Department For Environment, F., Office For National Statistics, 2017. Living Costs and Food Survey, 2015-2016. UK Data Service. doi:10.5255/ukda-sn-8210-4

Eftimov, T., Korošec, P., Koroušić Seljak, B., 2017. StandFood: Standardization of Foods Using a Semi-Automatic System for Classifying and Describing Foods According to FoodEx2. Nutrients 9. doi:10.3390/nu9060542

Fenland Technical Summary - MRC Epidemiology Unit [WWW Document], n.d. URL http://doi.org/10.22025/2017.10.101.00001 (accessed 6.28.19).

Green, R., Milner, J., Dangour, A.D., Haines, A., Chalabi, Z., Markandya, A., Spadaro, J., Wilkinson, P., 2015. The potential to reduce greenhouse gas emissions in the UK through healthy and realistic dietary change. Clim. Change 129, 253–265. doi:10.1007/s10584-015-1329-y

Hobbs, D.A., Lovegrove, J.A., Givens, D.I., 2015. The role of dairy products in sustainable diets: modelling nutritional adequacy, financial and environmental impacts. Proc. Nutr. Soc. 74. doi:10.1017/S0029665115003572

Horgan, G.W., Perrin, A., Whybrow, S., Macdiarmid, J.I., 2016. Achieving dietary recommendations and reducing greenhouse gas emissions: modelling diets to minimise the change from current intakes. Int. J. Behav. Nutr. Phys. Act. 13, 46. doi:10.1186/s12966-016-0370-1

Horton, P., 2017. We need radical change in how we produce and consume food. Food Sec. 9, 1323–1327. doi:10.1007/s12571-017-0740-9

Ingram, J., 2017. Perspective: Look beyond production. Nature 544, S17. doi:10.1038/544S17a

Institute of Medicine, 2014. Sustainable Diets: Food for Healthy People and a Healthy Planet: Workshop Summary, The National Academies Collection: Reports funded by National Institutes of Health. National Academies Press (US), Washington (DC). doi:10.17226/18578

Jerez, J.M., Molina, I., García-Laencina, P.J., Alba, E., Ribelles, N., Martín, M., Franco, L., 2010. Missing data imputation using statistical and machine learning methods in a real breast cancer problem. Artif Intell Med 50, 105–115. doi:10.1016/j.artmed.2010.05.002

Kearney, J., 2010. Food consumption trends and drivers. Philos. Trans. R. Soc. Lond. B, Biol. Sci. 365, 2793–2807. doi:10.1098/rstb.2010.0149

Laboratory, M.E.W., Research, N.S., 2019. National Diet and Nutrition Survey Years 1-9, 2008/09-2016/17. UK Data Service. doi:10.5255/ukda-sn-6533-13

Monsivais, P., Perrigue, M.M., Adams, S.L., Drewnowski, A., 2013. Measuring diet cost at the individual level: a comparison of three methods. Eur. J. Clin. Nutr. 67, 1220–1225. doi:10.1038/ejcn.2013.176

National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Food and Nutrition Board, Food Forum, 2019. Sustainable diets, food, and nutrition: proceedings of a workshop, The National Academies Collection: Reports funded by National Institutes of Health. National Academies Press (US), Washington (DC). doi:10.17226/25192

Oddy, D.J., 2003. From plain fare to fusion food: British diet from the 1890s to the 1990s. Boydell Press, Woodbridge, Suffolk.

Office For National Statistics, 2019. Living Costs and Food Survey, 2017-2018. UK Data Service. doi:10.5255/ukda-sn-8459-1

Orr, J.B., 1937. Food health and income: Report on a survey of adequacy of diet in relation to income. Macmillan and Company Limited,, London.

Perignon, M., Vieux, F., Soler, L.-G., Masset, G., Darmon, N., 2017. Improving diet sustainability through evolution of food choices: review of epidemiological studies on the environmental impact of diets. Nutr. Rev. 75, 2–17. doi:10.1093/nutrit/nuw043

Poore, J., Nemecek, T., 2018. Reducing food's environmental impacts through producers and consumers. Science 360, 987–992. doi:10.1126/science.aaq0216

Pye, S., Li, F.G.N., Price, J., Fais, B., 2017. Achieving net-zero emissions through the reframing of UK national targets in the post-Paris Agreement era. Nat. Energy 2, 17024. doi:10.1038/nenergy.2017.24

Reynolds, C., Hodgson, H., Bajzelj, B., 2019. An improved picture of the UK diet: Linking production, consumption and waste data to provide a better dietary picture. WRAP.

Reynolds, C.J., Horgan, G.W., Whybrow, S., Macdiarmid, J.I., 2019. Healthy and sustainable diets that meet greenhouse gas emission reduction targets and are affordable for different income groups in the UK. Public Health Nutr. 22, 1503–1517. doi:10.1017/S1368980018003774

Reynolds, C.J., Macdiarmid, J.I., Whybrow, S., Horgan, G., Kyle, J., 2015. Greenhouse gas emissions associated with sustainable diets in relation to climate change and health. Proc. Nutr. Soc. 74. doi:10.1017/S0029665115003985

Simons, L.P., Hampe, J.F., Guldemond, N.A., 2012. Designing Healthy Consumption Support: Mobile application use added to (e) Coach Solution. Bled eConference 34.

Sra, S., Nowozin, S., Wright, S.J. (Eds.), 2011. Optimization for machine learning. The MIT Press. doi:10.7551/mitpress/8996.001.0001

Stawarz, K., Cox, A.L., Blandford, A., 2015. Beyond Self-Tracking and Reminders: Designing Smartphone Apps That Support Habit Formation, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI' ' '15. Presented at the the 33rd Annual ACM Conference, ACM Press, New York, New York, USA, pp. 2653–2662. doi:10.1145/2702123.2702230

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., Collins, R., 2015. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 12, e1001779. doi:10.1371/journal.pmed.1001779

Tilman, D., Clark, M., 2014. Global diets link environmental sustainability and human health. Nature 515, 518–522. doi:10.1038/nature13959

Walker, P., Mason, R., Carrington, D., 2019. Theresa May commits to net zero UK carbon emissions by 2050 | Environment | The Guardian. The Guardian.

Willett, W., Rockström, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., Garnett, T., Tilman, D., DeClerck, F., Wood, A., Jonell, M., Clark, M., Gordon, L.J., Fanzo, J., Hawkes, C., Zurayk, R., Rivera, J.A., De Vries, W., Majele Sibanda, L., Afshin, A., Chaudhary, A., Herrero, M., Agustina, R., Branca, F., Lartey, A., Fan, S., Crona, B., Fox, E., Bignet, V., Troell, M., Lindahl, T., Singh, S., Cornell, S.E., Srinath Reddy, K., Narain, S., Nishtar, S., Murray, C.J.L., 2019. Food in the Anthropocene: the EAT-Lancet Commission on healthy diets from sustainable food systems. Lancet 393, 447–492. doi:10.1016/S0140-6736(18)31788-4

Wrieden, W.L., Leinonen, I., Barton, K.L., Halligan, J., Goffe, L., 2017. Is the UK diet sustainable? Assessing the environmental impact, cost and nutritional quality of household food purchases. Proc. Nutr. Soc. 76. doi:10.1017/S0029665117001811

Table 1. Meta-data overview of UK datasets of dietary assessments with multiple assessment moments and complete consumption/purchase information of at least one day.

Study name, institution, (reference)	**Study design**	**Sample moments**	**Target population**	**Recruitment**	**Sample size**	**Dietary assessment**			**Income recorded?**
						Method	**Portion size estimation**	**Administration method**	
EPIC Norfolk, University of Cambridge, (Day, et al, 1999)[1]	Cohort	1993-1998 (Phase I)	M/F, age 40-79 at recruitment	All HCP's in Norfolk were invited to participate. Participating HCP's (35) invited their all 40 to 79- year old clients to participate	30,283	24h-DR	Manual estimation. Suggestions on standardized portion sizes provided.	Pen and paper; self-administered; send by post	No
					25,348	FFQ (130 items)	Standardized portion sizes per item or food category. Participants recorded frequency of consumption per listing (9 choices ranging from never to 6 per day)	Pen and paper; personal interview; at test centre	"
					25,525	7-Day DD (cons.)	Free entry. Suggestions of standard portion sizes and pictures with example portion sizes provided. Days separated in 7 timeframes plus a field for snacks.	Pen and paper; supervised by nurse (day 1)/self-administered (other days); at testing site (day 1)/at home (other days)	"
		1994-1998	"	"	3,426	24h-DR	as previous 24h-	as previous	"

		18-month follow-up					DR	24h-DR	
					15,683	7-Day DD (cons.)	as previous DD	Pen and paper; self-administered; send by post	"
		1998-2000 (Phase 2) 3-year follow-up		"	NA	24h-DR	as previous 24h-DR	as previous 24h-DR	"
					11,449	7-Day DD (cons.)	as previous DD	as previous DD	"
					12,791	FFQ (130 items)	as previous FFQ	as previous FFQ	"
		2004-2011 (Phase 3) 13-year follow-up	"	"	6,702	7-Day DD (cons.)	as previous DD	as previous DD	"
					7,848	FFQ (130 items)	as previous FFQ	as previous FFQ	"
		2016-2018 (Phase 5) 23-year follow-up	"	"	1,665	24h-DR	Standardized portion sizes per item. Participants recorded number of standardized portions consumed.	Digital form; self-administered; at testing site	"
					8,416	FFQ (130 items)	as previous FFQ	as previous FFQ	"

EPIC Oxford, University of Oxford, (Davey, et al. 2003)	Cohort	1993-1999 (recruitment)	M/F, age 20+, large group of vegetarians	Invitations where send to (a) all 35 to 69-year-old listed with collaborating GP's in greater Manchester, Oxfordshire, Buckinghamshire. (b) members of vegan/vegetarian societies in the UK, aged 20+	57496 (regular meat eaters: 15,181; low meat eaters 7615; pescatarians: 7092, vegetarians: 15,426)	FFQ (130 items)	Standardized portion sizes per item or food category. Participants recorded frequency of consumption per listing (9 choices ranging from never to 6 per day)	Pen and paper; self-administered; send by post	No
		6 months after recruitment	"	"	30749	7-Day DD (cons.)	Free entry. Suggestions of standard portion sizes and pictures with example portion sizes provided. Days separated in 7 timeframes plus a field for snacks.	"	"
		5 years after recruitment (first follow-up)	"	"	38043	FFQ (88 items)	Standard portion sizes. Frequency of consumption (6 choices ranging from never to 6 per day or	"	"

							numeric entry)		
		2007 (second follow-up)	"	"	31695	FFQ (28 items)	None. Frequency of consumption (6 choices ranging from never to 6 per day or numeric entry)	"	"
					14309	7-Day DD (cons.)	as previous DD	"	"
		2010 (third follow-up)	"	"	32424	FFQ (113 items)	Standardized portion sizes per item or food category. Participants recorded frequency of consumption per listing (9 choices ranging from never to 6 per day)	"	"
Fenland Study, *Medical Research Council Epidemiology Unit*, *(Wareham et al.)*	Cohort	2005-2015 (phase 1)	M/F, born between 1950-1975, excluding pregnancy, diagnosed diabetes, inability to walk unaided,	People born between 1950-1975 and registered with the NHS in Cambridgeshire were invited to participate	12435	FFQ (130 items)	Standardized portion sizes per item. Participants recorded frequency of consumption per listing (9 choices ranging from	Pen and paper; self-administered ; at testing site	Yes

			psychotic or terminal illness				never to 6 per day)		
		2014-ongoing (phase 2, 4-year follow-up)	"	"	7000 (status May 2019; target 8000)	"	"	Digital form; self-administered; at testing site	"
					"	24h-DR	Standardized portion sizes per item. Participants recorded number of standardized portions consumed.	Pen and paper/digital form (transition July/Aug. 2015); self-administered; at testing site	"
UKBiobank, Cancer Epidemiology Unit, University of Oxford, (Sudlow et al., 2015)[1]	Cohort	2009/04-2010/09 (pilot)	GB, M/F, age 40-69 at recruitment (2006-2010)	NHS records were consulted to identify people within the age range and driving distance from one of the 22 assessment centres (total of 10 million were contacted)	70716	24h-DR	Standardized portion sizes per item. Participants recorded frequency of consumption per listing (9 choices ranging from never to 6 per day)	Online form; self-administered; at assessment centre	No
		2011/02-2011/04 (occasion 1)	"	"	100599	"	"	Digital form; unassisted; at home (link send by email)	"
		2011/06-2011/09	"	"	84265	"	"	"	"

		(occasion 2)							
		2011/10-2011/12 (occasion 3)	"	"	103792	"	"	"	"
		2012/04-2012/06 (occasion 4)	"	"	100248	"	"	"	"
UK Women's Cohort Study (UKWCS), University of Leeds, (Cade et al., 2017)	Cohort	1995-1998 (Phase 1)	Women who are vegetarian and non-red-meat-eating; 35-69 at time of recruitment (1995-1998)	Mail request to UK subscribers of cancer research fund and similar charities	35372 (5065 vegetarian, 4375 pescatarian)	FFQ (217 items)	Portions not specified for most items. Standardized portion sizes per item or food category for some items. Participants recorded frequency of consumption per listing (10 choices ranging from never to 6 per day)	Pen and paper; self-administered; at home (form send by post)	No
		1997-2000 (2-year follow-up)	"	"	1914	"	"	"	"
		1999-2002 (Phase 2; 4 years follow-up)	"	"	14172	FFQ (39 items)	Portions not specified. participants recorded frequency of consumption per listing (8 choices	"	"

							ranging from never to 6 per day)		
					12453	4-Day DD (cons.)	Recording weight/volume of foods as on package, as measured on a scale, or in standard household measures (last resort)	"	"
Family Food module of Living Cost and Food Survey (FFM-LCFS), ONS and DEFRA, (Department for Environment Food & Rural Affairs, 2018)	Cross-sectional	2008	UK households	Multi-stage stratified: Addresses clustered in PSUs, clustering postcodes sectors with < 500 households. 18 Households were sampled at random from 638 PSUs.	5845 households	14-Day food purchase diary (cons.)	None. Intake is estimated post-hoc from grocery purchases (receipts or manual diary entries), itemized meal descriptions when eating out and descriptions of free food and home-grown food	Pen and paper; self-administered; handed by the interviewer to be completed at home (free meals and home-grown food recorded in face-to-face survey)	Yes (separate face-to-face interview)
		2009	"	"	5825 households	"	"	"	"
		2010	"	"	5263 households	"	"	"	"
		2011	"	"	5692 households	"	"	"	"
		2012	"	"	5596	"	"	"	"

					households				
		2013	"	"	5144 households	"	"	"	"
		2014	"	"	5134 households	"	"	"	"
		2015	"	"	5080 households	"	"	"	"
		2016/2017	"	"	5020 households[2]	"	"	"	"
Kantar FMCG; Purchase Panel Kantar, Worldpanel Division (www.kantar.com; www.kantarworldpanel.com/global)	Cohort	Continuous	GB	Postal, social media, more direct methods for hard-to-recruit groups to establish a representative population based on Family Make-Up (taking into account age and number of different household members, including children), Social Class and Geographic Region	~30000 Households	Grocery purchases diary	None.	Scanning barcode on receipts using a handheld device; self-administered; self-administered; at home (recordings digitally transmitted to Kantar)	Yes (separate self-administered online form at recruitment and repeated after changes in household)
Kantar FMCG; Usage Panel; Kantar, Worldpanel Division (www.kantar.com; www.kantarworldpanel.com/global)	cohort	4 Times per year for one week each time continuous	GB, participants of the Kantar purchase panel	"	~11000 Individuals	7-Day household DD (cons.)	None. User selects the products used from a digital cupboard for each meal (allowing free	Online digital cupboard of purchased items; self-administered; at home (recordings	"

							entry), but do not specify consumed amounts.	digitally transmitted to Kantar)	
National Diet and Nutrition Survey (NDNS), NatCen/ NISRA and MRC Elsie Widdowson Laboratory, (MRC Elsie Widdowson Laboratory, NatCen Social Research., 2019)[1]	Cross-sectional	2008/2009, 2009/2010, 2010/2011, 2011/2012	UK, M/F, age 1.5+	Multi-stage stratified: Addresses clustered in PSUs, clustering postcodes sectors with < 500 households. 27 Households were sampled at random from 799 PSUs. In some only children were included. Additional recruitment was performed to achieve 100 participants in Northern Ireland, 100 in Wales and 200 in Scotland annually (out of ~1000 annual participants)	3450 adults, 3378 children	3 or 4-Day DD (non-cons.)	Free entry. (Descriptions of standardized portion sizes and pictures with example portion sizes provided)	Pen and paper; self-administered; at home (booklet provided by interviewer and collected after each recording day)	Yes (separate face-to-face interview)
		2012/2013-2013/2014	"	Multi-stage stratified: Addresses clustered in PSUs, clustering	1288 adults, 1258 children	"	"	"	"

				postcodes sectors with < 500 households. 28 Households were sampled from 323 PSUs. 10 Households sampled at random, 18 households were selected to have children					
		2014/2015-2015/2016	"	As above, but 316 PSU's	1417 adults, 1306 children	"	"	"	"
		2016/2017	"	As above, but 158 PSU's	494 adults, 454 children	"	"	"	"

“ = as above, -DR = 24h Dietary Recall, CEU = Cancer Epidemiology Unit, cons. = consecutive, 24h DEFRA = Department for Environment, Food & Rural Affairs, DD = Diet Diary, FFQ = Food Frequency Questionnaire, GP = General Practitioner, HCP = Health Care Practice, M/F = Males and Females, MRC = Medical Research Council, NISRA = Northern Ireland Statistics and Research Agency, ONS = Office for National Statistics, PSU = Primary Sampling Unit.

[1]No answer received on request to verify information.

[2]For this sample weight and measures of purchases where only recorded for 50% of the household diet expenses, whereas all households provided this information in other sample years.

All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.

This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

The authors declare no conflicts of interest.