



Department  
for Environment  
Food & Rural Affairs

## Privacy Impact Assessment: releasing Family Food Survey data (1974 - 2000) as open data

This assessment follows the template provided by the Information Commissioner's Office (ICO) in their publication [Conducting Privacy Impact Assessments: Code of Practice](#).

[Why are we undertaking a PIA?](#)

[Our information flows](#)

[Our consultation process](#)

[Our key privacy, compliance and corporate risks](#)

[Step five: Identify privacy solutions](#)

[Step six: sign off](#)

[Step seven: integrate PIA outcomes back into the project plan](#)

### Why are we undertaking a PIA?

Defra is planning to publish its archive of Family Food Survey data (1974 - 2000) as open data. It wants to do this in a way that ensures the privacy of individuals who have been surveyed throughout that time period is protected and respected. Defra needs to take care that data published as open data (a) does not contain any personal data and (b) cannot be reidentified. This Privacy Impact Assessment forms part of Defra's approach to managing risks associated with the publication of this archive.

Defra has committed to publishing **8,000 datasets as open data by June 2016 (#OpenDefra)**. While stimulating external and new uses of Defra data is a key goal of this project, it is also designed to kickstart cultural change across Defra and the transition to becoming a more collaborative, open organisation.

Since the launch of #OpenDefra, Defra has begun publishing a significant array of environmental and observational data, predominantly through its arms-length bodies.

It has developed an open data risk assessment (ODRA) process for working through risks associated with open data publication. To date, most data published could be described as 'low risk', with minimal issues relating to third party IP, personal data or commercial sensitivity.

The publication of Defra's archive of its National Food Survey data (1974 - 2000) as open data marks the first significant release of data that is derived from data provided by individuals. Since 2000, the National Food Survey has been retitled as the 'Family Food Survey' and is still administered by Defra in partnership with ONS. **'Family Food Survey' is used as an umbrella term for the data throughout this PIA.**

Since its origins, the Survey has captured a sample of households annually (approximately 7,000 from across the UK) and asks them to maintain a one week diary about what they eat and drink inside and outside the home. It's a fascinating snapshot of how British eating habits at home have changed over the years. While this PIA relates only to the publication of its data archive from 1974 - 2000, Defra has been collecting data as part of the Family Food Survey since 1940.

### **What are the goals from the publication of this data?**

The publication of Defra's archive of Family Food Survey data has benefits for people and organisations both inside and outside the department. The richness of information it provides about eating and drinking habits throughout Britain over decades could be useful for:

- analysing dietary and nutritional trends across Britain (for example, by combining it with nutritional data collected by the Food Standards Agency)
- understanding the relationship between dietary trends and other social forces that shape UK communities (employment, industrialisation)
- contributing to topical debates about obesity and sugar
- informing new teaching aides and visualisations in schools - an exciting way for students to engage with British eating habits over time.

Publishing its archive of Family Food Survey data also has benefits to Defra as a department:

- it's an engaging, easy to comprehend data source - a useful way to explore the possibilities of data, and how it reflects changes over time, with staff without specialist data backgrounds
- it encourages new approaches to publishing data within existing Defra professions - as part of this process, the statistics profession in Defra will develop valuable experience drafting privacy impact assessments, assessing disclosure control risks, documenting disclosure control mechanisms, working with data users, and developing anonymization skills.

The Family Food Survey has historically collected information not only about household eating and drinking habits across the UK, but also information about things like household numbers, ages of occupants, their gross income and type of employment. The need for a privacy impact assessment has been identified as part of the process preparing a version of the Family Food Survey data archive that is safe for release as open data. It will help Defra weigh the risks associated with the publication of this data in a way that balances the potential impacts of re-identification of individuals in the survey, as well as the public benefit in making this data more accessible. The task was to produce a treated dataset which was both valuable in its content to end users and which safeguards respondents' privacy to an acceptable level (both to Defra and them).

## **Our information flows**

### **How has this data been collected?**

Contributions to the Family Food Survey have been collected per household, via paper-based surveys, since 1939. Copies of the survey are distributed to a number of households across the UK. On average, approximately 7,000 households from across the UK are surveyed each year. Households are sampled at random, and linking food diary entries to particular households over time is not possible.

Some mechanisms around collection of the data have changed over the decades. In the 1970s, for example, surveys were typically supplied to the 'housewife' for completion, as person assumed in the household to have greatest understanding of what that household consumed over a two week period.

By the late 1980's, the designation 'housewife' was removed from the survey and it could be filled in by any household occupant. Food and drink consumed outside the home over a two week period were added as questions to the survey.

This survey, like all household surveys, would have been accompanied by a commitment to respondent confidentiality. The exact wording is now lost, but we continue to give this the highest priority in preparing the data archive for open publication.

Because this is a historic archive, information related to the original administration of the surveys has also been lost. The data Defra has access to may already have been cleansed of all direct identifiers, and had some disclosure control guidance applied. We do not know what other personal information may have been collected, and subsequently deleted, as part of data processing. We have commenced this PIA, and made changes to the data archive to reduce the risk of re-identification of individuals, based on the data that Defra has access to.

### **What does this data look like?**

The 'unprocessed' data archive available within Defra does not include any direct identifiers like name of survey recipient, address or national insurance number. However, the data does include a number of variables that may make re-identification of individuals and households in the data possible. The survey captures information including:

- Age of household occupants
- Gross income
- Employment type

- Whether any occupants were pregnant at the time of completion of the survey
- Whether occupants had certain benefits
- Sex

The data archive is currently accessible via the [UK Data Archive](#) to users who register and agree to the terms & conditions of its [End User Licence](#) (EUL). Due to the age of the data, it's unclear what disclosure control may have taken place prior to the uploading of Family Food Survey data to the UK Data Archive.

Defra would like to publish a version of the data archive more widely as open data. It's likely that some disclosure control will need to be undertaken prior to publication, to make the risk of re-identification of individuals contributing to the survey as remote as possible.

## **Our consultation process**

Defra has set up an internal project group to manage the publication of Family Food Survey data as open data, spanning representatives from the Food & Farming statistics unit, the Defra Data Programme and Communications Directorate.

The data archive doesn't include any contact details or direct identifiers for households who were involved in the survey. As such, contacting survey respondents individually to seek their feedback on the project is impossible.

As part of its preparation, the project group consulted:

- inside Defra, with the Information Rights Management team to understand the relationship between the Data Protection Act and anonymization of data.
- outside Defra, with the Office of National Statistics, to understand disclosure control guidance that is available and existing case studies of reuse of historic survey data, and with the UK Data Archive, to understand how the Family Food Survey data is managed via that platform.

The project group set up an external group of trusted testers, to explore five years of unprocessed test data (1974 - 1978) and provide feedback on variables that they would find most useful, as well as share their thoughts on potential privacy risks associated with the project. Feedback from testers has helped to shape Defra's own disclosure control testing.

A draft of this privacy impact assessment was also shared with the external testers, as well as the information rights team in Defra, to provide feedback on how it could be improved.

## Our key privacy, compliance and corporate risks

### Risk One: Individuals can be re-identified using the Family Food Survey data

The Information Commissioner's Office has [made clear](#) that data protection law does not apply to data that has been anonymised; that is, changed in such a way that the data subject is longer identifiable. Defra is following ONS disclosure control guidance and the ICO Anonymization Code of Practice to reduce the likelihood as much as possible that this data could be re-identified. Nonetheless, we need to think through the impacts to individuals, and to Defra, were re-identification to occur.

In some circumstances, even if identification of an individual from the data is possible, there may be no negative impact. An individual might recognise their own household from a diary entry. A neighbour or associate might make an educated guess about the identity of a household in the data. The food diary results do not divulge anything sensitive or unusual, and/or the recognition may occur privately. We can foresee instances where even if identification occurs, there's no impact on individuals or government.

There is still a risk that, even in the event no harm is experienced as a result of re-identification, Defra may be found to have breached its obligations under the Data Protection Act. The ICO may launch an investigation into the circumstances which resulted in the re-identification of individuals. A finding by the ICO that Defra has negligently disclosed personal data about individuals could result in civil penalties and a fine of up to £500 000. The ICO has stated that '**where there is evidence of re-identification taking place, with a risk of harm to individuals**, the Information Commissioner will be likely to take regulatory action' (20, Anonymization Code of Practice).

Finally, there is a risk that a household or households is identified from the Family Food Survey data, with negative outcomes. A combination of certain unique household information is found to be that of a well-known household - e.g. a Minister, an MP, a celebrity, a public intellectual, athlete etc. The details of these food diaries form the basis of media stories revealing the eating habits of these public figures. These eating habits may not be unusual, or they may reveal details that are embarrassing to the figures. Alternatively, a food diary entry may be so unusual as to incentivise re-identification of the household surveyed, and - if achieved - result in a public focus on that household's eating habits.

## Solution

Defra has taken steps to ensure that the likelihood of identification is as remote as possible<sup>1</sup>. In order to assess the likelihood of identification, the team:

- ran a range of count queries, looking for unique and low combinations of results in the data. Due to the sample nature of the survey, a large return of unique and low combinations did not necessarily mean the nature of the household information captured was unique (i.e. that only one household of four, with two children and two adults, from Bath, was captured by the survey does not mean that such a household is rare in Bath);
- applied Skinner and Elliot's data intruder simulation (DIS) method<sup>2</sup> on a variety of combinations of variables to assess potential disclosure risk; and
- spoke with external testers about their experiences of the data, and combinations of variables they considered to be unique.

To reduce the risk of identification as much as possible, Defra has decided to:

- Remove data records for households with more than 10 members
- Where ages are recorded, band them into 7 age groups
- Delete the gross income of head of household
- Band the net family income into equivalised deciles (within the dataset, not national equivalised decile income points)
- Merge the variables recording receipt of various benefits (Family Credit, Income Support, Housing Benefit, Working Family Tax Credit) into a single binary variable
- Delete variables on religion (these were only recorded in Northern Ireland)
- Not publish the 'person' table which recorded a number of details on each member of the household (age, sex occupation etc). A new person table will contain two binary variables identifying whether the household contains a) a pregnancy and b) any children under 3.

As part of the data release, Defra will provide a mechanism for people to alert the team to any instances of identification or other privacy-related issues in the data. Text will be added to the data release pages with a contact email address ([familyfood@defra.gsi.gov.uk](mailto:familyfood@defra.gsi.gov.uk)) asking data users to alert the team to any issues. If a privacy-related issue is raised and found to be valid, the data will be removed from data.gov.uk and any other locations until the issue is resolved.

---

<sup>1</sup> In a case between the Department of Health and the Information Commissioner [2011] EWHC 1430 (Admin), the court found that: '**the risk of identification must be greater than remote and reasonably likely for information to be classed as personal data**' .

<sup>2</sup> Skinner, C. J., & Elliot, M. J. (2002). A measure of disclosure risk for microdata. Journal of the Royal Statistical Society: series B (statistical methodology), 64(4), 855-867

**Risk Two: Regardless of any actual identification of individuals from the Family Food Survey data, the public perceives it to be a breach of their trust in government.**

There are increasing sensitivities around how government collects, accesses and uses our personal data. There may be a general perceived risk that Defra has not been managing its Family Food Survey data properly and securely (absent of any incidents, or based on identification of households). Participants in future surveys may, as part of the fallout from such a risk, refuse to participate in the Family Food Survey/other government surveys.

Defra is heavily reliant on the continuance of surveys to collect data. As such, maintenance of the response rate into the future needs to be taken into account. Care must be taken to not only protect privacy, but be perceived as protecting privacy, in a way that also shows that the data is being used for the benefit of the nation, while respecting those who volunteered their time and data to DEFRA and its predecessors in the past.

The ongoing administration of the Family Food Survey is the responsibility of the Defra statistics profession, who are bound by the [UK Statistics Authority Code of Practice](#). The Code of Practice in its current form didn't exist at the time this historic data was collected (1974). The statistics profession inside Defra could be found to be in breach of the existing Code of Practice.

**Solution**

We will publish our privacy impact assessment alongside the publication of this data archive, to show people the steps taken prior to publishing this data as open data. Demonstrating a rigorous risk process, and being open and transparent about internal consultation, changes to the data and risks considered, can be powerful mechanisms for engendering trust in government.

We will also work with the Defra communications directorate to prepare a set of reactive messages in the event concerns are raised about Defra's collection and publication of the survey results. Importantly, this release should be positioned as a positive, people-friendly data publication - an opportunity to explore food and drink trends around Britain from the 1970s - 2000.

**Step six: sign off**

*Who approves the privacy risks involved in the project?*

For this project, a data governance group has been assembled comprising:

- the senior responsible officer for data within Defra
- the Head of the Statistics Profession in Defra
- The Director for Food & Farming and

- the Director General for Evidence

This group must review the PIA, discuss it with the Family Food Survey team and approve publication of the data archive as it recommends.

Defra will look to establish a permanent sub-committee of the Data Governance Board - an internal board helping to implement the wider Defra Data Programme - to review future releases that may involve some personal data or other risk.

### **Step seven: integrate PIA outcomes back into the project plan**

The Food and Trade statistics team owns this data and the Food and Trade Statistician is the current asset owner. The Food and Trade Statistician has responsibility to ensure publication conforms with the recommendations of the overall assessment as signed off by the data governance group.

### **Preparing the next tranche of data (2001 - present) for release**

This PIA concerns the preparation of Family Food data from 1974 - 2000 for release. Following the 18th February, the project team will begin preparing for the release of data to the present day, collected as part of the Office of National Statistics (ONS) Living Costs & Food Survey. As the data from 2001 to present is jointly administered by ONS and Defra, a different governance process will be required to approve sign off of the data for publication.

Defra expects that the more recent data may require additional disclosure control, due to its age and the increasing availability of public information sources (e.g. social media) that may increase the risk of identification. Defra will begin with the data specification agreed on for its 1974 - 2000 release, and conduct a penetration test of the more recent data after it has been processed according to this specification. Defra will also continue to engage with external users and provide updates to changes to the data set for publication as a result of this process.