



Home Office

## **Publishing detailed datasets in Immigration Statistics**

Produced by the Migration Statistics team

[MigrationStatsEnquiries@homeoffice.gov.uk](mailto:MigrationStatsEnquiries@homeoffice.gov.uk)

First published: 22 August 2019

# 1 Background

---

- 1.1 This document summarises Home Office consideration of the format of the data published as part of the *Immigration Statistics* series, and how the Home Office reached decisions about publication.
- 1.2 In undertaking this work the Home Office has taken care to ensure that the risk of identifying individuals or releasing personal or sensitive data is minimal and proportionate.
- 1.3 The Home Office collects a range of data as part of the administrative processes required to manage and progress immigration casework. While the current [Immigration Statistics](#) publication provides an overview of these processes, the current format has limitations on the amount of data that can be published in a clear and accessible way, which is conducive to robust and meaningful analysis. To address this, the Home Office considered alternative ways to publish data on immigration and to develop more detailed migration data tables that provide greater flexibility for users of the statistics.
- 1.4 It is essential to have a sound understanding of the impacts and behaviours of those who come to the UK, in order to ensure policy and operational decisions are based on a robust evidence base, which in turn ensures the immigration system runs fairly and efficiently. This includes understanding the personal characteristics of those interacting with the immigration system. Having a robust evidence base is an essential part of this and is in the public interest, but it is important to consider the value of the statistics against any risks to individuals when producing the statistics.
- 1.5 The publication of detailed migration data tables will have a number of benefits for users.
  - **Increase trust in the statistics** by presenting in an open data style, users can interrogate the data as required, increasing confidence that the data are being produced in an impartial manner.
  - **Enable better understanding of the data** by setting it out clearly and simply, while giving users the tools to explore the data in more depth.
  - **Increase the value to a range of users** by publishing ‘summary tables’ which allow users to find key, headline statistics, as well as more detailed data for specialist users.
  - **Improve the quality of the statistics** by streamlining the production process, allowing producers to focus on the accuracy, reliability, and coherence of the statistics.
  - **Modernising the data** by publishing the data in a format that is compatible with a range of specialist software. This in turn will allow for better analysis and presentation of the data through the use of analytical and data visualisation software.

## 2. Disclosure control

---

- 2.1 Whilst there is additional value to users in publishing the statistics in this format, and it is considered to be for the public good, the decision to publish has also been assessed in line with the [GSS/GSR Disclosure control guidance for tables produced from administrative sources](#). That involved considering both the risk (“the likelihood of an individual, business or household or other statistical unit (or related attributes) being identified in a published table”), and the impact (“the impact that disclosure of information will have on the individual, business or household concerned”) of doing so.
- 2.2 The number of people with certain combinations of variables (such as nationality, age and sex) can be low; as a result there will be occasions where individuals may be identifiable within the data. While this risk cannot be completely eliminated without significantly reducing the value of the data, the Home Office have taken a number of steps to reduce the risk and impact of such disclosure.
- 2.3 Nationality and age are not physical characteristics. This means that an individual of a certain appearance cannot easily be linked to the data based on these characteristics alone, without prior knowledge about the specific age, nationality, or other circumstances of such an individual. This reduces the likelihood of any disclosure impacting any individual.
- 2.4 The majority of the published data relate to the UK as a whole. This high level of geographical aggregation reduces the disclosure risk. Where more granular geographical data are provided, these are not combined with other personal information (such as age, sex, or nationality) unless necessary. In cases where personal characteristics are provided alongside more granular geographical data, relevant disclosure controls will be put in place.
- 2.5 Where it is in the public interest to publish statistics containing ‘special category’ data, as set out in the ‘General Data Protection Regulation’ (GDPR), these are not combined with other personal information (such as age, sex or nationality) unless necessary. In exceptional circumstances, where there is value in combining ‘special category’ data with other personal characteristics, relevant disclosure controls, such as suppression, are put in place to ensure individuals cannot be identified. For example, see the data published in [asylum claims on the basis of sexual orientation](#).
- 2.6 Consideration was given to using suppression more widely across the data. However the negative impact this would have on the use and accessibility of the data was considered too high alongside the relatively low risk and impact to the individual.
- 2.7 Where possible, the Home Office have grouped variables into broad categories. For example, ages have been grouped. These groupings have been considered alongside policy and other interests in specific groups to ensure the data continue to add value, while reducing any risk of disclosure. For example, there is a particular interest in the number of 16-17 year olds claiming asylum, as these cases are more likely to lead to age dispute cases. This narrow age group has been included in a summary table (which is not combined with other personal characteristics) to

minimise the risk of disclosure. When combine with other personal characteristics, the broader 'under 18' category is used.

- 2.8 In some cases, individuals may appear in multiple datasets (for example, asylum applications, initial decisions, and appeals). Where possible, variables have been grouped to minimise the risk of identifying individuals by linking them between datasets, however, any risk cannot be eliminated completely without diminishing the value of the data. However, as there are no physical or special characteristics within these datasets (with the exception of sex), it would be difficult to identify an individual without prior knowledge of their circumstances.
- 2.9 Overall, while there is a risk that some individuals may be identified in the data, it is considered that the impact of this is low and appropriate controls have been put in place to minimise the risk, while maintaining the value of the data.

### 3 Assessing privacy risk

3.1 The Privacy Impact Assessment (PIA) screening questions are a list of questions to consider at the beginning of any project involving personal data. They shaped the project and informed the decision that a full Privacy Impact Assessment is not necessary.

**Table A: Privacy Impact Assessment Screening Questionnaire**

	Question	Yes	No	N/A
Q1	Will the policy involve the collection of new information about individuals?		X	
Q2	Will the project compel individuals to provide information about themselves?		X	
Q3	Will information about individuals be disclosed to organisations or people who have not previously had routine access to the information?		X	
Q4	Are you using information about individuals for a purpose it is not currently used for, or in a way it is not currently used?		X	
Q5	Does the project involve you using new technology which might be perceived as being privacy intrusive? For example, the use of biometrics or facial recognition.		X	
Q6	Will the project result in you making decisions or taking action against individuals in ways which have a significant impact on them?		X	
Q7	Is the information about individuals of a kind particularly likely to raise privacy concerns or expectations? For example, health records, criminal records or other information that people would consider to be particularly private.		X	
Q8	Will the project require you to contact individuals in ways which they may find intrusive?		X	

3.2 Table A above provides an example of how the data in Immigration Statistics has been assessed on the basis of the PIA screening. It is anticipated that the majority of the data released in this project will follow this example. On this basis it has been decided that a PIA is not required and this document instead sets out the approach to suitably anonymise the data, however it may be the case that future releases will require a PIA.

3.3 A key consideration of the development phase of this project was how to increase data availability whilst being mindful of the potential sensitive nature of the data. Following discussions with data experts, and with due regard to the Data Protection Act 1998, it was agreed that the most appropriate way to open up more of the data was to publish standalone datasets that address particular topics or themes.