

IC-244085-D1T2

Thank you for your letter dated 14 August, in which you have asked us to consider our response to FOI2023/28161 and IR2023/36246, and the use of section 44 (1) (a) exemption.

A reminder of the original request FOI2023/28161 we received:

“Thank you for your response. These requests have now been ongoing for quite some time, the original request made 20th October 2022.

I refer back to your response with your reference FOI2023/11634 where you state:

“The requested breakdowns required multiple datasets to be joined to obtain the information and, as these datasets are structured differently and contain different information, this created complexities when attempting to combine the datasets. As a result of this, bringing these datasets into alignment was not possible within the cost limit.”

Please provide these datasets. I have the skills required to process these datasets. To be clear, I want these datasets in any computer readable format. I do not want a response telling me to use Stat-Xplore.”

This was followed by an Internal Review IR2023/36246:

“Thank you for your response.

I am not happy with the response and therefore I am here requesting an internal review.

To be clear, my request should be accepted with the delivery of anonymised datasets.”

For the FOI and subsequent review, we said the information was held but cited an exemption under section 44 (1) (a): Prohibition on disclosure.

We will address the points raised in your letter however to help explain our position we felt it best to describe the process the department goes through to produce the datasets, and how we usually use them to help respond to FOI requests.

Administrative Data Process

The datasets used for analysis are derived from snapshots of the raw data collected on millions of individuals by each of the computer systems used by the department to administer the different benefits available. For each benefit there is a usually a separate system used, e.g., Jobseeker’s Allowance Payment System (JSAPS) is used for administration of most aspects of JSA. For some benefit processes, such as

when sanctions are considered, the Decision Making and Appeals System (DMAS), the Decision Makers and Appeals Case Recorder (DMACR) system and the Labour Market System (LMS) are also used. There is also a Customer Information System (CIS) which contains personal individual level records such as date of birth, and address for those who interact with the DWP, and this information is supplemented with death registration information regularly received from the General Register Office.

Data extracts are taken from the many different administrative systems used to manage and process claims and, depending on the age of the systems, some are taken as quarterly snapshots of the data held, whilst others are updated overnight. These extracts contain millions of records and are used to produce internal management information (MI) and official statistics. The extracts used for MI are usually compiled as volumetrics, so it is not possible to match between the various extracts produced. However, the files made available to analysts to produce official statistics, are provided at the individual customer level thus enabling data matching, robust quality assurance and analysis to facilitate the publication of a large suite of [official statistics](#).

Internal security issues

Technical staff who provide the files to those producing MI and Statistics have appropriate internal security clearance to handle this data and must ensure that any files passed on to other staff within the department are appropriately protected.

In this case, the customer level files produced are pseudonymised before they are made available to analysts. Data that has undergone pseudonymisation remains personal data and is in scope of data protection law. This means that access to these files is restricted to named analysts who have current business case approval to use the data.

External users are prohibited from accessing any of these files unless they are contracted by the department to perform agreed analysis or research and, should this be the case, access would only be permitted on-site under strict access conditions.

SAS Dataset handling

The primary analytical tool used in the department is [SAS](#), as it is capable of processing and matching the millions of cases the department processes every month. This means that the files made available to analysts are usually provided on internal servers as either SAS data views or SAS data sets. Analysts use SAS to interrogate these files, perform the necessary cleaning, quality assurance and data matching needed to build the full customer level journey information required to produce high quality summary datasets which are then published as official statistics either on [Stat-Xplore](#), or via separate ods summary tables (spreadsheets) on the [DWP Statistics website](#).

For each set of statistics, there is a respective 'background, information and methodology' (BIM) document published which gives details of the different source systems used, and the issues which may have arisen when compiling the data to produce the statistics. Sanctions statistics uses several different data sources and so is particularly more complex to compile.

The following links provide the information relevant for this case:

[DWP benefits statistical summaries: policies and statements - GOV.UK](http://www.gov.uk/government/statistics/dwp-benefits-statistical-summaries-policies-and-statements)
(www.gov.uk)

[Universal Credit statistics: background information and methodology - GOV.UK](http://www.gov.uk/government/statistics/universal-credit-statistics-background-information-and-methodology)
(www.gov.uk)

[Benefit sanctions statistics: background information and methodology - GOV.UK](http://www.gov.uk/government/statistics/benefit-sanctions-statistics-background-information-and-methodology)
(www.gov.uk)

[Universal Credit sanctions statistics: background information and methodology - GOV.UK](http://www.gov.uk/government/statistics/universal-credit-sanctions-statistics-background-information-and-methodology) (www.gov.uk)

[Sanctions durations and rate: background information and methodology - GOV.UK](http://www.gov.uk/government/statistics/sanctions-durations-and-rate-background-information-and-methodology)
(www.gov.uk)

Fol request handling

When an Fol request is made and the information is not already published we will look at the analytical SAS datasets to see if the information is held there, and if so, extract the relevant data to produce the requested information. In some cases, it may be necessary to select and merge several SAS datasets to obtain the information requested.

As the information required can be derived from different administration systems, which are constructed in different ways, some difficulties can arise when trying to match the relevant data, and this makes it harder to obtain the requested information within the Fol cost limit. See the relevant BIM documents above for further details.

As detailed in the BIM, to select all the relevant cases and perform the correct matching process, complex SAS programs are used by analysts to take account of the various data sources and different periodicities of the data extracts. Even with these programs, some information is not easy to compile, and additional iterative programming may be required to attempt to ascertain the information.

In some cases, where data may not be available in the analytical SAS files a request might be needed to extract additional information from the primary source system to include in the SAS file. Due to the specification of each data extract, in most instances this process would cause the Fol cost threshold to be exceeded.

Information requested

Prior to this request, the requestor originally asked for information on caseloads, sanctions, and deaths, and we did look at using the analytical SAS datasets available to us to compile the information. However, due to the complexities of how

the information is recorded, particularly with respect to sanctions, the work needed to collate and extract the information required would have exceeded the FoI cost limit. This cost exemption was supported by your recent decision notice [IC-228829-N0M7](#).

After explaining these difficulties, and despite making an alternative suggestion that we could use a different methodology to provide some relevant information for those on UC, the requestor just asked to receive copies of the all the datasets we mentioned so they could attempt to compile the information from those files themselves.

As part of our evidence for IC-228829-N0M7, we explained that in total 63 datafiles would need to be processed to compile the information requested. However, for this request we have now identified that a total of 83 SAS datasets would be required. The number of files for each type of file is summarised as follows:

Caseload files: UCLS/UCFS (1), JSA (19), ESA (19), IS (19)
Sanction decisions files: UCLS (1), UCFS (1), JSA (1), ESA (1), IS (1)
Deaths file: (1)
Legacy Benefit Person caseload file (used for geography): (19)

As all these files contain individual customer level data and some have several cases per customer and are also updated regularly, they range in size from several thousand rows to nearly 300 million rows. For example, in July 2023 there were 6.1 million people on UC, but as the UC caseload file contains all historic UC information and is updated each month, it currently contains nearly 300 million rows of data. In addition to the security issues already mentioned, and the sheer volume of data involved, we are unable to provide you with a copy of this withheld information at this time. However, to assist with the case we have provided details of the contents of the datafiles in the attached table.

Section 44 (1) (a): Prohibition on disclosure

You asked:

“The Commissioner notes that DWP has confirmed that it considers that section 123 of the SSAA 1992 prohibits disclosure of the requested information.

- A. Please explain why DWP considers that the datasets would identify individuals.
- B. Is it DWP’s position that the entirety of the information in the datasets would identify individuals?
- C. Can the datasets be redacted or use <5 for small numbers to prevent identification?”

We consider that the information is exempt due to section 44 (1) (a), specifically that we are prohibited by section 123 of the SSAA 1992 to release data which identifies individuals. The SAS datasets and data views contain individual level data which has been pseudonymised, meaning that the files contain personal information. In

addition, there are other variables in each file which could also potentially allow individuals to be identified. The table supplied gives details of these variables.

This pseudonymisation is necessary for secure handling of the files and to allow analysts to perform any necessary data matching of individuals within and between datasets in preparation of statistics. We believe that the removal of this pseudonymisation would render the datasets of no practical use, as no data matching would be possible. Therefore, we believe this data remains personal data and we do not have the lawful basis to share.

You also wrote:

“The Commissioner notes that section 123(3) of the SSAA1992 provides two conditions in which the disclosure of this kind of information by DWP would not constitute an offence. These are that:

- The information in question has previously been disclosed to the public with lawful authority
- If the information in question is disclosed in the form of a summary or collection of information so framed as not to enable information relating to any particular person to be ascertained from it.

D. Please explain why DWP has concluded that the gateways to disclosure or exceptions to the prohibition contained within the enactment are not engaged in respect of this request.”

Due to the security and disclosure issues provided above, the specific datasets requested have not been previously disclosed to the public. However, the department has made information held in those datasets available “in the form of a summary or collection of information so framed as not to enable information relating to any particular person to be ascertained from it” via the very large number of official statistics routinely published on [Stat-Xplore](#) and the [DWP Statistics website](#).

Finding a way forward

The statement we made regarding the difficulties in matching the required information was based on our original interpretation of the request which we admitted may have been incorrect. We put forward an alternative suggestion in our responses to FOI2023/11634 and FOI2023/28161 and this was also mentioned again in the decision notice for IC-228829-NOM7.

We suggested using our existing sanctions rate methodology to provide some information on UCFS caseloads and sanctions and deaths within 3 months. Using this method would allow us to produce a spreadsheet ods file which would contain the following summary information, broken down by region:

For the count date in April 2019, and every 3 months after that,

- (a) the number of people on UCFS in the sanctionable conditionality groups,

- (b) the number of people in a) who died in the following 3 months,
- (c) number of people on UCFS receiving a sanction,
- (d) the number of people in c) who died in the following 3 months.

We believe that this information may be of use to the requestor.

We hope this provides suitable information to help you with your investigation. If you have any further questions regarding this response, please get in contact.

Please note that we ask for all information submitted to you as part of this ICO investigation to be deleted / returned to us on conclusion of your investigation.

DWP response for IC-244085-D1T2 – Details of Datasets requested

Dataset	Coverage requested	No. of files	No. of variables	Number of rows	Personal/ identifiable variables	Code required to ID the correct cases
UCFS and UCLS caseload	Second Thursday of the month from 11 Jan 2018 to 14 Jul 2022	1	17	Latest file (subject to retrospection, full UCFS/LS caseload): 290 million After narrowing coverage: 90 million	Encrypted NINO, DOB, age, sex, location	Yes, code required to extract relevant period and conditionality
JSA caseload	Each quarter end from Nov 2017 to Feb 2022	19	31	Nov 2017: 410,000 Feb 2022: 110,000	Encrypted NINO, location, asylum seeker flag	No
ESA caseload	Each quarter end from Nov 2017 to Feb 2022	19	44	Nov 2017: 2.3 million Feb 2022: 1.7 million	Encrypted NINO, disability, location, asylum seeker flag	Yes, code required to extract claimants with the correct level of benefit to be sanctionable
IS caseload	Each quarter end from Nov 2017 to Feb 2022	19	38	Nov 2017: 570,000 Feb 2022: 190,000	Encrypted NINO, location	No
UCLS sanction decisions	From Jan 2018 to Jul 2022	1	38	One file for full caseload (like UC caseload): 920,000 After narrowing coverage: 84,000	Encrypted NINO, DOB, DOD, age, sex, location	Yes, code required to extract relevant period and decision outcome (adverse vs non-adverse)
UCFS sanction decisions	From Jan 2018 to Jul 2022	1	43	One file for full caseload (like UC caseload): 1.3 million After narrowing coverage: 1.3 million	Encrypted NINO, DOB, age, sex, location	Yes, code required to extract relevant period
JSA sanction decisions	From 1 Dec 2017 to 28 Feb 2022	1	42	One file for full caseload (like UC caseload): 15.3 million After narrowing coverage: 23,000	Encrypted NINO, DOB, age, sex, disability flag, ethnicity, lone parent flag, location	Yes, code required to extract relevant period and decision outcome (adverse vs non-adverse)

DWP response for IC-244085-D1T2 – Details of Datasets requested

Dataset	Coverage requested	No. of files	No. of variables	Number of rows	Personal/ identifiable variables	Code required to ID the correct cases
ESA sanction decisions	From 1 Dec 2017 to 28 Feb 2022	1	37	One file for full caseload (like UC caseload): 490,000 After narrowing coverage: 2,200	Encrypted NINO, DOB, age, sex, disability flag, ICD 10 code, ethnicity, lone parent flag, location	Yes, code required to extract relevant period and decision outcome (adverse vs non-adverse)
IS sanction decisions	From 1 Dec 2017 to 28 Feb 2022	1	35	One file for full caseload (like UC caseload): 32,000 After narrowing coverage: 13,000	Encrypted NINO, DOB, age, sex, disability flag, ethnicity, lone parent flag, location	Yes, code required to extract relevant period and decision outcome (adverse vs non-adverse)
Deaths	From 1 Dec 2017 to 28 Feb 2022 (and to Jul 2022)	1	6	Not available for full file. After narrowing coverage: 3.3 million	Encrypted NINO, DOB, DOD, sex	Yes, code required to extract deaths in the relevant period
Legacy Benefit Person caseload (used for geography)	From 1 Dec 2017 to 28 Feb 2022	19	34	Nov 2017: 17.6 million Feb 2022: 15.8 million	Encrypted NINO, DOB, age, sex, ethnicity, date reached SPA, DOB of youngest child, age of youngest child, location	Yes, code required to extract recipients of relevant benefits