

## Equality Data Asset: Proof of Concept

### Purpose

1. This paper describes ONS plans and progress to date on developing an Equality Data Asset (EDA) in partnership with the Cabinet Office. The EDA is part of a broader programme of work, the Equality Data Programme, and is intended to meet Inclusive Data Taskforce (IDTF) recommendations in relation to improving the granularity of available data for exploring equalities and enabling intersectional and longitudinal analysis.

### Timing

2. The Equality Data Asset (EDA) Proof of Concept (POC) dataset is currently hosted on the [Secure Research Service \(SRS\)](#)<sup>1</sup>, with access initially restricted to relevant ONS and Cabinet Office staff only. Work on creating the first iteration of the EDA in ONS's [Integrated Data Service \(IDS\)](#)<sup>2</sup> environment is due to begin this financial year. This will start by linking [Companies House People with Significant Control register](#) (PSC) data to the 2021 Census. Any learning from the POC dataset will need to inform this process. In addition, to allow time for acquisition of data for subsequent iterations of the EDA, there is an urgent need to identify the priorities for additional data sources.

### Recommendations

3. Members are invited to:
  - i. note the existence of the EDA Proof of Concept dataset and suggest any relevant research priorities that it might be used to address;
  - ii. provide feedback on the design of the POC dataset and suggest any improvements to inform the development of the EDA in the IDS;
  - iii. suggest any priorities for additional data to be included in future iterations of the EDA.

### Background

4. The [Equality Data Programme \(EDP\)](#) was announced by the UK Government in December 2020. Its aim is to improve the data and evidence available to understand the barriers that people from different backgrounds face across the UK. In early 2021, ONS partnered with the Cabinet Office to deliver the EDP.
5. During its first phase, ONS carried out analysis on existing ONS data in line with Cabinet Office requirements. This aimed to identify the personal characteristics and circumstances associated with poorer outcomes in relation

---

<sup>1</sup> The Office for National Statistics (ONS) Secure Research Service (SRS) is a Trusted Research Environment (TRE), which gives accredited or approved researchers secure access to a wealth of de-identified, unpublished data to work on research projects for the public good.

<sup>2</sup> The Integrated Data Service (IDS) is a cross-government service, for which the Office for National Statistics is the lead delivery partner, which enables co-ordinated access to a range of high-quality data – critical to informing policy decisions and improving public services.

to wealth, mortality, crime, life satisfaction and digital exclusion. The [findings from the analysis](#) were published on 30 January.

6. The second phase of the EDP involves the development of an Equality Data Asset (EDA).
7. The vision for the EDA is that it will link together relevant datasets to enable the comparison of outcomes across multiple areas of life for the same people and over time. These areas of life are:
  - i. Work and pay
  - ii. Education and skills
  - iii. Health
  - iv. Justice
  - v. Living standards
  - vi. Family, community and life satisfaction
8. The intention is that the EDA will be a resource, not just for government researchers, but for the wider research community, including regional decision makers, academics and third sector organisations.
9. To realise its full potential, the EDA will need to provide data at a sufficient level of granularity to enable meaningful analysis of even the smallest population groups.
10. The aim is to take an iterative approach to developing the EDA, bringing onboard more data sources in successive iterations, and for development to be undertaken by ONS's Integrated Data Service (IDS).
11. Building on the success of the SRS, the IDS will provide secure and safe access to accredited researchers to datasets from a range of sources across government departments and the devolved administrations, to enable vital research and decision making in the UK. As such, it is the right strategic vehicle for the development of the EDA.
12. ONS began work on developing the POC dataset in 2022. This has recently been completed and data made available to Cabinet Office staff in the SRS to test its utility and inform further development.
13. The EDP remains a priority for the Women and Equalities Select Committee (WESC) who recently held the Minister for Equalities to account for the commitments made in relation to it.

## Discussion

14. The Equality Data Asset Proof of Concept (POC) dataset links a snapshot of the Companies House People with Significant Control (PSC) register (as at October 2021) to the 2011 Census (see **Annex A** for details of the PSC information included in the linked POC dataset).
15. This linkage was chosen as a starting point for the POC for several reasons:
  - i. As the most complete record of the population and its personal characteristics, the logical starting point for the EDA is Census data; Census 2021 data were not available when work was begun
  - ii. The PSC data are publicly available, so their use avoided any delays in securing access to the data to carry out the data linkage
  - iii. It goes some way towards addressing a known data gap in the information about the characteristics of business owners, identified as a priority for the

All-Party Parliamentary Group for Black and Minority Ethnic Business Owners among others.

16. Although Companies House data provides one of the most complete records of all businesses, one of its limitations is that it does not include many small or independent businesses. ONS has estimated as many as 52% of all businesses may not be registered with Companies House. Nevertheless, it is one of the few available business data sources with the necessary information to enable linkage to Census data.
17. Analysis of the quality of the linkage work indicated that the linkage was good, with an estimated 98% of matched records being a true match. Although almost 3 million PSC records were matched to a Census record, this represents only 30% of PSC records. Work is now beginning to explore any biases that may have arisen as a result and explore solutions to address them.
18. Any learning from the development and analysis of the POC will be used to inform the ongoing development of the EDA. Beyond recreating the PSC-Census linked dataset in the IDS using Census 2021 data, ONS and Cabinet Office will soon be beginning work to consider the additional data sources to ideally bring into the EDA and in what order of preference.
19. As noted in NSIDAC(23)05, the paper on subnational statistics, achieving the vision of the EDA is not without its challenges. These include securing access to the data, overcoming any issues with linking them and ensuring that any linked data are not disclosive. However, these are challenges that are already being considered as part of the ongoing development of the IDS.

## **Conclusion**

20. This paper has described our progress towards developing an Equality Data Asset to meet IDTF recommendations. Members are invited to provide feedback on the plans and progress and suggest any priorities for research using the POC dataset and any suggested improvements in the design of the EDA in the next phase. In addition, Members are asked to propose priorities for future iterations of the EDA.

**Paola Serafino, Centre for Equalities and Inclusion, Statistical Coherence and Inclusion Division, ONS**  
**20 April 2023**

## **List of Annexes**

**Annex A Information from the Companies House People with Significant Control register contained within the linked POC dataset**

## **Annex A Information from the Companies House People with Significant Control register contained within the linked POC dataset**

Note: To protect the confidentiality of those included in the linked dataset, any information that could identify an individual, such as name, address, company name, etc., has been removed and replaced with pseudonymised identification numbers.

- Pseudonymised unique person identification number
- Pseudonymised company identification number
- Date that PSC ceased to be a person of significant control
- Date that PSC notified their person of significant control status
- Category of submitted accounts information. Describes the type of accounts filed.
- Accounting reference date, i.e. the financial period.
- The made up date of the latest accounts annual return filed. If no return has been filed, this will be left blank.
- The date the next accounts annual return is due to be filed.
- Type of company
- The current status of the company
- Where company was originally registered
- Date company was dissolved, where relevant.
- The date that the company was incorporated.
- Number of General Partners.
- Number of Limited Partners.
- Total number of charges registered against the company
- Total Number of Outstanding charges registered against the company
- Total Number of Part Satisfied charges registered against the company
- Total Number of Fully Satisfied charges registered against the company
- The made-up date of the latest annual return filed. If no return has been filed, this will be left blank.
- The date the next annual return is due to be filed.
- Type of business activity conducted by the company.