

## Written evidence from the UK Statistics Authority<sup>1</sup> (TEB30)

### Public Administration and Constitutional Affairs Select Committee Transforming the UK's Evidence Base inquiry

Following the submission of the Office for National Statistics' (ONS) written evidence to the Committee's *Transforming the UK's Evidence Base* on 31<sup>st</sup> August 2023, I then gave evidence to the Committee on 5<sup>th</sup> September 2023. I am now able to provide some supplementary evidence, as requested, on several topics of interest.

#### **The Integrated Data Service (IDS)**

As you will be aware, the IDS is a cross-government project, for which the ONS is the lead delivery partner. The project is a key enabler of the National Data Strategy and seeks to securely enable coordinated access to a range of high-quality data assets built, linked and maintained for richer analysis. Please find below some further detail on the background of this project and the progress towards its delivery.

#### What is the scope of the IDS?

The scope of the IDS is to deliver a secure scalable modern data service which operates on a cloud-native platform, hosting a rich and diverse data catalogue consisting of indexed and linkable data with the latest provision of data science and generative AI potential. The service has been designed to better inform effective policy making.

The vision of the IDS is to address the lack of a central integration platform that can cater for the future needs of both data providers and analysts looking to utilise integrated data to develop cross-cutting analytical results. The IDS builds on the success of the Secure Research Service (SRS) and offers to significantly reduce the time it takes to negotiate and access data and the provision of data assets.

The IDS provides a secure environment that enables streamlined data sharing across government improving the ways that data are made available via cloud native technologies, modernising the way departments and their professionals operate. The IDS is the first of its kind in the UK and will be setting the precedence for how data is being processed on a cloud native platform.

#### When is it expected to be delivered?

The programme has been in development by the ONS over the last 18 months and is funded until March 2025 (under the current Spending Review). After this date, the IDS becomes a live running service.

---

<sup>1</sup> Submitted by Professor Sir Ian Diamond, National Statistician, UK Statistics Authority

### What is the cost of the programme?

The programme secured funding from HM Treasury (HMT) until the end of the investment period (financial year March 2024/25). The cost of the programme is estimated to be £228.7m which covers the development and running costs from 2020 – 2025. Furthermore, the programme continues to assess funding options beyond March 2024/25.

### Who are the users likely to be?

The IDS is designed for use by accredited analysts, within government and the wider research community. The ambition for the IDS is to have every government analyst, roughly estimated at 14,000 individuals, capable of utilising the platform to better inform decisions for the public good.

### What data do you expect to be available on the service?

There are currently 81 datasets available in the IDS from across government. This includes high-value data assets, such as levelling up; climate change and net zero. Additionally, health data assets are underway with identified datasets being indexed by the Reference Data Management Framework (RDMF) – which enables multiple data to be linked and analysed, creating new comprehensive data assets – and published on the IDS so that analysts can link data according to their requirements.

The programme intends to continue to work with data owners across government and the private sector to acquire more datasets in conjunction with the RDMF. However, this is dependent on data owners signing up to data sharing agreements to make this data available.

In accordance with the Central Digital and Data Office's roadmap for 2022-25<sup>2</sup>, departments have agreed to share their essential shared data assets across government, including through IDS. This further enables the IDS as a Trusted Research Environment to facilitate and support this commitment.

However, discussions with government analysts have highlighted a range of concerns about how current incentives for departmental data sharing fit with the needs of ministerial-facing departments. There is also a wider financial risk regarding other department's ability to fund activity such as data cleansing, which may limit their ability to effectively share data. Although HMT set out the expectation that OGDs will support data sharing in all SR21 settlements, no specific funding was provided, which may limit activity in some cases. As part of the IDS Programme, ONS is working with Chief Data Officers across government to minimise frictions around the sharing of data via IDS. One of the pilots in development is looking at Data Ownership and Stewardship approaches to streamline the governance arrangements and make it quicker for departments to agree to share data via IDS, and for analysts to subsequently access that data for a broad range of

---

<sup>2</sup> <https://www.gov.uk/government/publications/roadmap-for-digital-and-data-2022-to-2025/transforming-for-a-digital-future-2022-to-2025-roadmap-for-digital-and-data>

analysis in the public good. As always, I would welcome support from the Committee to share and promote the benefits of data sharing across government for the public good.

#### What safeguards will be in place to protect data?

The IDS is a trusted research environment which means it adheres to the 5 Safes in accordance with the Digital Economy Act (DEA); The 5 safes of secure data are as follows:

- Safe projects - Is this use of the data appropriate, lawful, and ethical?
- Safe people - Can the users be trusted to use it in an appropriate manner?
- Safe settings - Does the access facility limit unauthorised use?
- Safe data - Is there a disclosure risk in the data?
- Safe outputs - Are the statistical results non-disclosive?

These principles enable the safeguards and governance for the IDS to operate with sensitive data which in turn ensures public confidence in the security and processing of data. Access to the IDS platform is granted via secure gateway in line with the data legislation; furthermore, the IDS utilises strict policies around the cleaning, linkage, validation and controlling data.

The IDS Programme is also working across ONS in the development of key governance through policy creations that will enable safeguards and the appropriate use of data. The policy workstream, which is coordinated by ONS' Data Governance, Legislation and Policy and Security and Information Management teams, is helping to develop adequate governance for the programme via policy development. In developing safeguards, the programme employs the following principles:

- Adapting successful policies within the ONS and across government analytical communities (e.g., GSS, GSR, GES) that can support the programme.
- Working with the National Statistician's Data Ethics Advisory Committee, which is underpinned by the UK Statistics Authority's (UKSA) ethics framework for the use of data for statistical, research and analytical purposes, to identify and mitigate any potential ethical risks at project-level.
- Access to all data are controlled through the concept of a analytical 'project', within supporting business and technical processes linked to user need.
- An overarching programme Data Protection Impact Assessment (DPIA) is maintained to define key activities and associated data risks. Continued engagement with the Information Commissioner's Office on the DPIA as it is maintained and updated as the programme develops.

The programme also adheres to the UK Statistics Authority/ONS Data Protection Policy (required by the Data Protection Act 2018 and the General Data Protection Regulation).

#### **The ONS website**

The Committee also asked for some insight into the current condition of the ONS website

and any plans to change the site in the future. Below I have outlined our vision for dissemination, of which our website is an integral part, as well as some exploratory work we are undertaking to see how we could use AI technology to address some of the challenges with our existing website.

### Our Vision for Dissemination

The ONS website supports the Statistics for the Public Good strategy by helping to build trust in evidence, enhance understanding of social, economic and environmental matters and improve the clarity and coherence of our communication. By helping people to be aware of the ONS and to find, understand and explore our data, statistics and analysis we are giving people the information they need to make decisions, and act, at a national, local and individual level.

Our vision for statistics dissemination goes beyond the website. We want people to have trust in our data and analysis. We know that our users want to find trusted ONS information wherever they look – whether that's on the ONS website, on social media, in the media or through search engines. Our users want ONS answers to their questions and we are exploring a range of different approaches to serve this need, including providing answers to questions using Large Language Models (LLMs).

Our goal is for users to understand our data, statistics and analysis more quickly and easily, with the right contextual information to help people know how they can use them. We want our users to explore and tailor our information so they can find what is important to them - whether that is by creating their own datasets based on ONS data or through our expert curated view of key insights for the economy or society.

Our priorities for the website in recent years have been delivering the capability to support census 2021 outputs and the reliability of the service to all our users, particularly in response to the additional demand for ONS data on the economy, in response to changes in the cost of living. We're currently running a package of work to address and improve website performance to meet demand and our next priority will be programmatic access to our data via application programming interfaces (APIs). This will improve the agility of all users of our data, both internal and external, to consume and gain insights from the ONS website.

We have also focused on improved search both on the ONS website and through greater visibility of our data and insight in search engines and in the media.

This year we are also setting the future direction for how we create and manage our statistical content in a more efficient and structured way to enable business agility and flexibility for our users, aligned to their broad range of needs. This will set out a forward plan to transform ONS data and insight and will make the case for the additional funding needed to deliver on our ambitions.

### StatsChat

Additionally, the ONS Data Science Campus are currently exploring how new tools and technology can help the organisation disseminate information more effectively. We have developed a new product, 'StatsChat', that uses LLMs to search and summarise text from across our website, and present relevant sections of our web pages to user's natural language questions.

We are aiming to make this available to a small selection of users for testing and fine-tuning, so that we can improve the relevance of the responses and provide assurance from a data ethics, data protection and security perspective.

### **Stakeholder engagement**

The ONS conducts a wide range of user and stakeholder insights, consultations and listening exercises. This engagement is essential as it provides us with actionable insights on users' and stakeholders' views on the strength of their relationship with the ONS, feedback on its outputs, and on how stakeholders access and use our statistics and analysis.

As part of this, the ONS's Engagement Hub conducts annual stakeholder 'deep dive' research and an annual stakeholder satisfaction survey. I understand the Committee is interested in understanding more about these exercises and insights from recent examples.

The deep dive research is conducted through in-depth interviews with senior representatives from around 45 key stakeholder organisations. The stakeholder satisfaction survey is an online questionnaire aimed at a wider range of users from a variety of sectors and roles to provide broader insight. Deep dive participants include those from central and local government departments, devolved administrations, research institutes, think tanks, public bodies such as NHS England and the ICO, international partners, business representative bodies and charities. The stakeholder satisfaction survey reaches similar types of organisations, with a wider range of responses at senior manager, operational, public affairs, analyst, researcher, policy maker and economist levels.

Deep dive interviews took place in summer 2022 and the findings were positive. Many stakeholders said that the organisation had built on and maintained its reputation for independence, trustworthiness, quality and reliability. They also felt that the ONS had developed its reputation for being flexible, agile and responsive to changing needs. Additionally, the ONS was seen to be working more collaboratively with policymakers than it had in the past.

The stakeholder satisfaction survey was conducted in early 2023. It found respondents to be positive across key sentiment measures on trust, quality, and on the ONS producing statistics which are relevant to issues of the day. There were also positive views expressed about the ONS as an organisation with reliability, responsiveness, and willingness to help being cited. It was also noted that ONS staff were knowledgeable and helpful.

There were areas highlighted for improvement in both the stakeholder deep dive and satisfaction survey. These included how the ONS works with both devolved governments and heads of the statistical profession in government departments; improving the ease of finding the right people to speak to in the organisation; and more regular, strategic overviews of the ONS's work (for stakeholders to be able to connect different topics better). Some participants referenced a need for further scrutiny to understand some data anomalies which had occurred in mid-2022.

These findings are shared throughout the ONS, including with the National Statistician's Executive Group, and are used to inform planning and prioritisation. We have implemented measures to respond to the issues raised as part of a wider programme of ongoing external affairs improvements, which we continue to monitor with further research.

The ONS conducted a subsequent stakeholder deep dive in autumn 2023 and are currently analysing the findings. The latest ONS annual stakeholder satisfaction survey is currently live and will be open for responses until 22 January 2024<sup>3</sup>.

### **Full business case on population and migration statistics improvements**

As you are already aware, next year I will be making a recommendation to Government on the future of the population and migration statistics system in England and Wales. I understand that the Committee has requested some additional detail surrounding the financial aspects of this transformational work.

In the outline business case for the Future of Population and Migration Statistics programme, initial cost estimates of a potential census in 2031 range from £1.3 billion to £2 billion, with increases expected across all phases of such an operation.

The ONS is working to produce a full business case (FBC) for our proposals to improve our population and migration statistics. The FBC will be developed in the context of the forthcoming recommendation to UK Government, and the response from Government. At this stage, while the recommendation remains in development, it is difficult to provide an accurate updated estimate of cost.

The FBC is expected by HM Treasury in late 2024. We will be able to provide the Committee with further information on costs at a later date.

### **Migration statistics**

As part of improving population statistics we are also transforming international migration statistics. Our latest estimates<sup>4</sup>, year to June 2023 are official statistics in development and are provisional. We revised our June 2022 and December 2022 estimates upwards due to a combination of more data and methodological improvements.

---

<sup>3</sup> <https://consultations.ons.gov.uk/external-affairs/stakeholder-satisfaction-survey-2023/>

<sup>4</sup> <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration>

International migration estimates are produced using three key sources: Home Office border data linked to a person's travel visa for non-EU nationals, which made up 82% of total immigration in 2023; tax and benefit data (known as RAPID) for EU nationals; and International Passenger Survey data for British nationals. We are most confident with Home Office border data and have an ambition to produce all migration statistics from these data in future.

We work very closely with Home Office to procure and use border data linked with visa data to produce migration estimates. The ability of free movement for British nationals and some EU and non-EU nationals makes the current method a challenge for those that don't require visas. However, there is further data held by the Home Office, known as Advanced Passenger Information, that would help with our research, particularly for British nationals. We have requested these data and would like to see Home Office accelerate this request.

Census 2021 data confirmed our position that the administrative data we use for non-British nationals is robust and that the international passenger survey data does not measure actual migration patterns well due to people changing their intentions.

Rather than rebasing once a decade, following a decennial census, to correct for any drift in our population estimates, we aim to produce statistics that do not 'drift' from the truth. Our Dynamic Population Model based population statistics show how drift in both population and migration statistics can be mitigated<sup>5</sup>. That does not remove the need to revise estimates as the data and methods mature.

Long-term international migration uses the UN definition of a migrant, that is someone that changes their country of residence for 12 months or more. To produce timely estimates, we therefore have to make assumptions based on previous behaviour. As more time passes, we are able to update those assumptions with data of actual travel. We therefore become more confident in our estimates over time. For example, our June 2022 estimates now have complete data to show if a migrant has stayed or left for 12 months and we therefore have less uncertainty around those estimates compared to the provisional June 2023 estimates.

We have recently published experimental uncertainty measures<sup>6</sup> for our admin-data based migration estimates for the first time. These show our users how our confidence increases once we have complete data that meet the required definition.

We also described the nature of provisional estimates that are subsequently revised and the reasons behind these revisions. This was picked up and presented accurately in the media and in playing back conversations with our core users. The Office for Statistics Regulation (OSR) recently published a review of their recommendations on migration statistics<sup>7</sup>. The OSR considered we sufficiently described uncertainty to our users,

---

<sup>5</sup><https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/methodologies/dynamicpopulationmodelimprovementstodatasourcesandmethodologyforlocalauthoritiesenglandandwales2021to2022>

<sup>6</sup><https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/bulletins/longterminternationalmigrationprovisional/yearendingjune2023>

<sup>7</sup> <https://osr.statisticsauthority.gov.uk/publication/office-for-national-statistics-transformation-of-long-term->

although we recognise these are experimental and will continue to update our users as they develop.

I hope that you find this additional information useful. Please do let us know if we can assist the Committee further on any of the issues discussed in this letter, or with any of its other inquiries.

*December 2023*