

CONFERENCE OF EUROPEAN STATISTICIANS

Workshop on Statistical Data Collection

WP.2-1

10-12 October 2017, Ottawa, Canada

22 August 2017

Transformation of UK Business Data Collection

Debra Prestwood, Nick Barford & Louise Morris (Office for National Statistics, United Kingdom)

debra.prestwood@ons.gov.uk, nick.barford@ons.gov.uk

Abstract

The UK Office for National Statistics (ONS) is in the process of a major transformation of its technology infrastructure. The drivers for the transformation are to improve data quality and statistical outputs; reduce burden on respondents; extend the range of analysis possible from linked datasets across economic and social data domains; and to deliver efficiency savings.

This involves the implementation of a new Enterprise Architecture to facilitate use of generic platforms, services and processes across the statistical outputs of the ONS, including economic statistics, social statistics and Census.

From the business (and social) data collection perspective this involves moving surveys online; rationalising surveys and harmonising the definitions of variables collected; and integrating the use of non-survey data. The ONS has been constrained in using administrative data sources until very recently. The passing of the Digital Economy Act 2017 makes provisions for data sharing.

The ONS is interested in other countries' experience of these challenges.

UNECE Workshop on Statistical Data Collection 10-12 October 2017, Ottawa

Paper title: Transformation of UK Business Data Collection, Debra Prestwood, Office for National Statistics, UK

Session: Modernisation of data collection

Executive Summary

The UK Office for National Statistics (ONS) is in the process of a major transformation of its technology infrastructure. The drivers for the transformation are to improve data quality and statistical outputs; reduce burden on respondents; extend the range of analysis possible from linked datasets across economic and social data domains; and to deliver efficiency savings.

This involves the implementation of a new Enterprise Architecture to facilitate use of generic platforms, services and processes across the statistical outputs of the ONS, including economic statistics, social statistics and Census.

From the business (and social) data collection perspective this involves moving surveys online; rationalising surveys and harmonising the definitions of variables collected; and integrating the use of non-survey data¹. The ONS has been constrained in using administrative data sources until very recently. The passing of the Digital Economy Act 2017 makes provisions for data sharing.

The ONS is interested in other countries' experience of these challenges.

Introduction

The United Kingdom Office for National Statistics (ONS) is partway through a major transformation. The aims of the transformation are to embrace new technology to enable use of non-survey data to improve statistical outputs and analysis, improve efficiency and reduce respondent burden.

This entails modernisation of its technology infrastructure, implementing a new Enterprise Architecture, replacing IT processing systems and developing new statistical methods. A Service Orientated Architecture approach is being taken to the design and build of new software processing systems across ONS, with the principle of 'develop once, use many times' for the production of statistical outputs across the domains of Business, Social and Census data.

This paper describes how ONS is moving our suite of 80 business surveys online, rationalising those surveys and integrating administrative and other non-survey data sources. To facilitate this, a new Statistical Business Register is also being developed.

¹ In practice this could be administrative sources, commercial or other 'Big data' sources

Background

Around 80 surveys are conducted by the ONS to businesses in Great Britain and the UK. These underpin the production of the UK's economic statistics and compilation of the National Accounts. They range in frequency from monthly, to quarterly to annual surveys. They have been set up over time in response to user need including various EU statistical legislation; although some rationalization of surveys has been done over time this has been on a small scale, dependent on available funding.

In October 2015 ONS was allocated additional funding to transform, including transformation of data collection for business, social and Census. In particular, the aims are to move to online collection and make use of non-survey data in place of surveys, where possible. Up until very recently the latter has been constrained by the lack of appropriate UK legislation to enable data sharing of administrative data sources collected by the different areas of the UK government.

In April 2017 the Digital Economy Act (DEA) was passed which removes this constraint. From January 2018 ONS expects to have access to new administrative datasets for the first time.

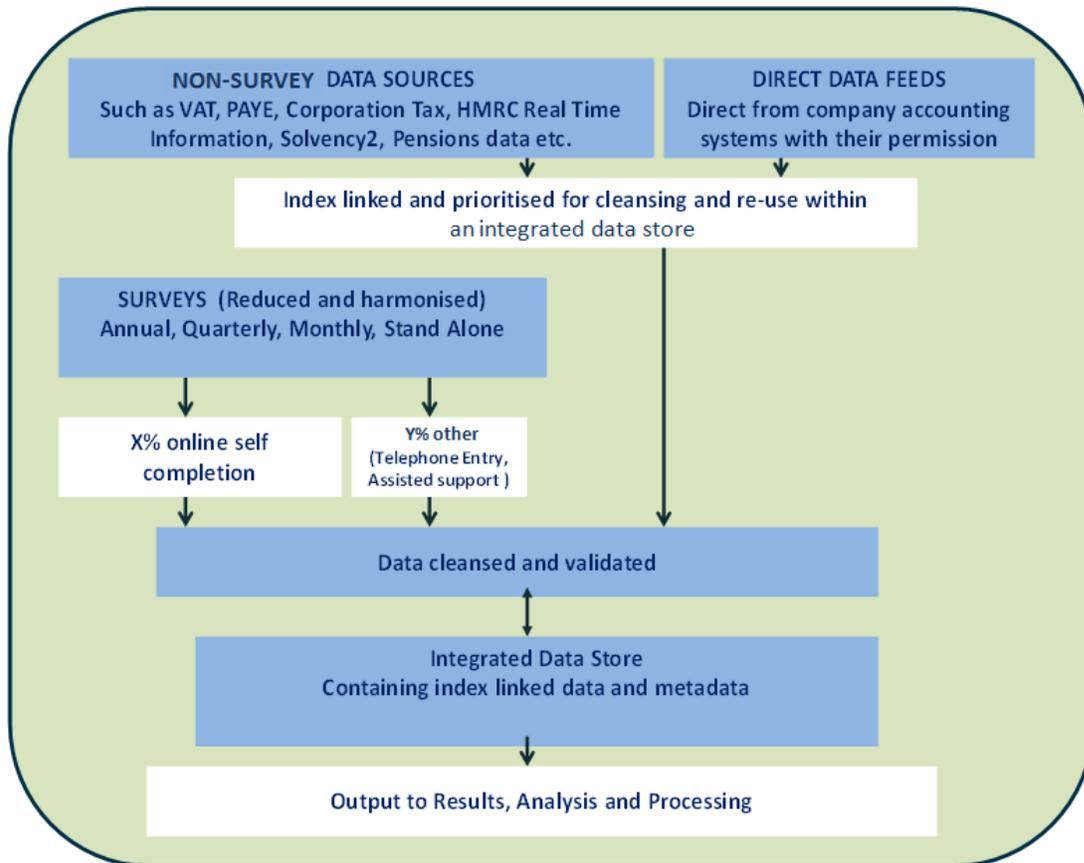
Vision for Business Data Collection

- Administrative and other non-survey data are the preferred sources for the production of statistics, with surveys being used for model calibration or to fill gaps;
- Any remaining business surveys are largely rationalised and modularised ie combined to reduce the number of times a respondent is separately surveyed.
- Survey data collection is digital by default (online). In future, ONS online data collection services are used across UK government departments;
- Integrated data sources are accessible from a linked data store to support ONS outputs – so that data processing, analysis and publication make the best use of all data sources;
- Collection operations are supported by common services and IT systems across ONS business, social surveys and Census.

Where surveys are still required, they will have been rationalised from various paper surveys to online collection on a monthly, quarterly and annual basis. Where possible, the definitions of the concepts being measured, for example turnover, will have been harmonised across surveys. A business will therefore receive an online questionnaire on a monthly/quarterly/annual basis, split into 'modules'. A module will contain questions on for example, turnover or employment, rather than collect these data on separate paper surveys. There would also be perhaps half a dozen other stand-alone business surveys to collect data on specific issues that do not neatly fit the modular structure.

A small proportion of survey responses are still expected to be collected via Assisted Digital means and by telephone. The vision is then to integrate the non-survey and survey data by linking it at individual business level through registers, for use in production of statistical outputs and to enable enhanced analysis.

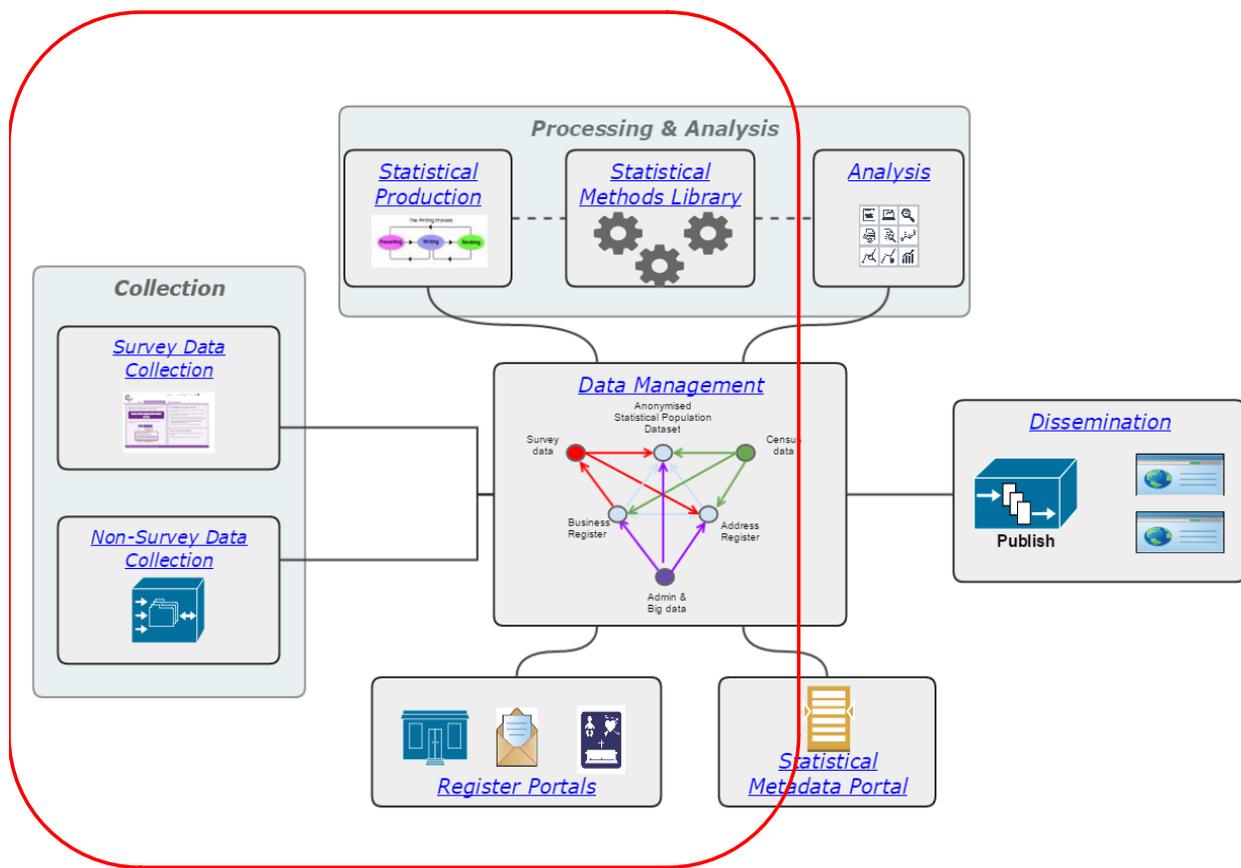
Business Data Collection Future State



Enterprise Architecture

ONS has developed an ONS Enterprise Architecture (EA). This is based on the [Generic Statistical Business Process Model \(GSBPM\)](#), the internationally agreed standard for statistical processing developed by the United Nations Economic Commission for Europe (UNECE). The ONS EA will be realised as a small number of service oriented processing platforms, each comprising a number of products to realise business functionality. The red boundary in the following diagram indicates where data collection activities are impacted.

Enterprise Architecture Platforms (Release 11)



A Platform is a shared IT capability consisting both of a related set of technology infrastructure and applications, and the development and operations specialists that build, enhance, and support it. The core products from a collection perspective are:

- Survey Data Collection platform – a secure registration and transmission portal for respondents to complete and submit surveys online via electronic questionnaires (eQs);
- Non-Survey Data Collection Platform – a secure collection environment for administrative data, commercial data, ‘Big data’ and other non-survey data sources to be submitted to ONS;
- Data Management platform – a secure access controlled area where all data (survey, non-survey, unit and aggregate level) will be stored;
- Statistical Production platform – where processes will be orchestrated to perform editing, imputation for non-response, weighting etc;
- Statistical Methods Library – a store of the methods required for statistical production across the GSBPM;
- Statistical Metadata portal – metadata will be key to the new overall approach.

- Register portals – the user interface to enable usual register functions to be carried out using the platforms described above.

Approach to New Systems Development

The new digital services and platforms are being developed in-house within ONS, contracting in technical development resources. Agile delivery approaches are being used ie develop small increments quickly, review and continue to develop. The approach is to develop the ‘Minimum Viable Product’ (MVP) of the platforms and services to meet business goals, then develop these MVPs further by adding more functionality to meet the next set of user needs.

Increased statistical methodology resources are required to develop the new statistical methods and processing associated with the use of non-survey data. The aim is also to make use of generic statistical methods as far as possible, stored within the Statistical Methods Library. The majority of statistical methods development is being done in-house, with some contracting out of this work where ONS does not have the capability or would benefit from an external perspective.

Approach to Business Surveys Statistical Redesign

A number of ‘activities’ have been defined to indicate the extent of transformation. Surveys could be subject to one of the following or a combination²:

- *Translation*: Where a paper survey is translated online ‘as is’ ;
- *Rationalisation*: Where surveys are effectively combined with others when moved online and data definitions are harmonised;
- *Integration*: Where surveys are moved online and reduced in sample size and/or, variables collected, given the use of non-survey data.

The existing 80 business surveys have been categorised into eight ‘Survey Groups’, for example Short Term Surveys, Financial Surveys etc. These reflect surveys that have synergies and are likely to be rationalised/integrated.

A ‘roadmap’ for the transformation of surveys has been developed taking into account ONS’ priorities for improving economic statistics and National Accounts, using non-survey data sources such as Value Added Tax (VAT) turnover supplied by businesses as part of their government tax returns, administrative returns regarding pay to employees (known as Pay As You Earn (PAYE)) and other sources such as commercial data on financial transactions. The current priority is to transform the monthly surveys

² In practice a survey could be translated online, then rationalised with others, then integrated. Alternatively rationalisation and integration could be implemented simultaneously.

measuring turnover and employment using VAT and PAYE sources. These are also the most costly surveys.

Further priorities are redesigning the financial and balance sheet surveys to take account of commercial data sources in order to improve measurement of the financial flows between the economic sectors following the financial crisis of 2008. And using PAYE real time information to improve the sampling frame for annual estimates of hours and earnings.

Development of a new Statistical Business Register

A new statistical business register is being developed, as well as a new legal unit based list of the business population, called the business index. The new Business Index and Statistical Business Register (SBR) will replace the current Inter-Departmental Business Register (IDBR).

A first step in the register process is the creation of legal units, and these can have a wider use than just the SBR. To enable this a new product called the Business Index (BI) is being trialled: utilising new data sources, this will provide a complete list of business legal entities in the UK for the first time. This will primarily be used by the wider government community, to see the legal view of a business, which in many cases is more relevant than the economic view provided by the enterprise view but also serves as the first step in the business register process. The BI will then feed these legal units into the new SBR system, where the family of statistical units will be created.

SBR Functions

The SBR is being designed to provide the following improvements:

Frame function : The new SBR will provide a frame which facilitates mixed data acquisitions, including both traditional sample survey collection and the use of administrative data. The frame methodology is still being designed, but the solution has to enable the population to be partitioned. Firstly to facilitate the use of administrative data by providing a population of units; this will then be linked to administrative data to append appropriate variables. Secondly, for the remaining part of the population to be processed through a subsequent sampling and estimation procedure.

The sampling functionality will be built to accommodate sampling at any level of statistical unit , i.e. enterprise group, enterprise or local unit, which will improve the service available.

Spine function : It will also provide a spine function enabling linking of business datasets and facilitating the use of administrative and survey data, to create statistical outputs.

By linking the business registers with address and people spines, it is possible to develop new analytical outputs and enhance the utility of data. For example, knowing that a person works at a certain business, can then be linked to characteristics of that business to generate information. Secondly the spine function will enable the linking of administrative data and facilitate the production of statistics. For

example monthly VAT turnover can be linked to the spine, and converted into statistical units, with better industrial classification, enabling the production of outputs.

Analytical function: For many years there has been growing demand to use the business register for economic analysis of the business population. The IDBR has a number of shortcomings due to its original design being primarily as a sampling frame. These mostly concern the under coverage of units, age and metadata for auxiliary variables. The most important of these is clear metadata and time-stamping for each variable. Secondly to always use the most up to date data – the previous system employed rules which in some circumstances maintained older data from better quality sources. The new system will hold different versions of key variables such as turnover and employment. The most up to date data will be available and its data provenance very clear.

Expanded population: The IDBR is known to have under-coverage of very small entities. This was managed through estimation using aggregate data from administrative sources, but the wider role of registers requires this under-coverage to be addressed. ONS has also recently benefitted from new data sharing provisions (Digital Economy Act 2107), which will enable access to important new data sources to fill this gap.

The main new data source will be Self Assessment Tax. This will be used to identify very small sole traders and partnerships. There are a large number of businesses in this group, but by definition their total output is only a small part of the economic output of the UK.

Direct Data Feeds from Businesses

The focus of this is to work collaboratively with businesses and their primary software/systems suppliers to scope the range of potential information that could be obtained directly from businesses' accounting, analytical and other systems. The aim being that acquiring business data in this way could also minimise the burden on businesses and ONS by circumventing the survey process.

Receiving data in this way direct from businesses' systems would flow in through the Non-Survey Data Collection platform, as in the case of other non-survey data sources.

Research is required into the following questions:

- i. Whether company accounting, data systems and electronic records can directly deliver a range of information;
- ii. Exploration of the businesses' view of how we should best access standard accounting and other key information as opposed to using survey media for some series/variable;
- iii. Examination of any software barriers to such a process with key business software providers;
- iv. Drawing on the experience of other NSIs who collect data directly from business systems in this way ; as well as the views/experience of commercial information collectors from business organisations

- v. To ascertain whether direct data collection and online survey collection strategies could flow through a common platform covering all direct (company) as well as survey data collections;
- vi. Whether these survey and direct strategies could be based around a standardised delivery and transfer process.

Progress so far

Moving online

ONS has invested in developing an online survey management system tailored to improving the respondent experience. Businesses can log in securely using a single username and password to see the complete list of surveys for which they have been selected. The system also allows a survey questionnaire to be allocated to another member of the organisation for completion. ONS staff can monitor the progress of survey completion and extract extensive metrics on the rate of completion of eQs, when during the day/week are using the system etc. The development has included extensive testing with respondents to improve the design of the electronic questionnaires (eQs) and this is continuing as each survey is moved online to continually improve the respondent experience.

At the time of writing of this paper, five business surveys are fully online:

- Monthly Commodity Inquiry;
- Monthly Retail Sales Inquiry;
- Monthly Wages and Salaries Survey;
- Quarterly Business Survey
- biannual UK Innovation Survey.

This equates to approximately 58,000 businesses completing surveys online. Currently approximately 1.4m paper questionnaires are completed by businesses across the 80 business surveys. A further eleven surveys are using the new survey management system to securely submit data in electronic file format to ONS. A platform upgrade is planned for November 2017 which will allow further surveys to be moved online; the next tranche is planned for January 2018 and includes the Monthly Business Survey with a sample size of 32,000.

Until now, surveys have been ‘translated’ online ie the paper questionnaire has been adapted for online completion but ‘as is’ in terms of definitions and variables collected. That is, no rationalization of surveys has yet been launched or sample sizes reduced to take account of the use of non-survey data sources. The approach being taken is that once non-survey sources are available (under the new legislation described above) and research into methods completed to allow multi-source estimation of aggregate estimates using both the survey and non-survey data sources, the emphasis will shift to putting transformed surveys online.

Decision points on the readiness of surveys to move online in the transformed state form part of the overall roadmap for business surveys transformation. The fallback position will be that we continue to translate surveys online with transformation research continuing in parallel until methods and new platforms are ready.

Transformation of surveys

Methodological research is underway into the redesign of the Monthly Business Survey, starting with the retail sector as the priority. This survey measures turnover and employment activity on a monthly and quarterly basis respectively. The approach being taken is to revisit the 'output requirement' ie what is required in terms of measurement by the users. This will feed into how the variables are harmonized and collected in the modular way described above.

Research is in progress with respondents to understand what data they are able to provide and how easily. Also to analyse the VAT data to understand it's comparability with the survey unit level data and develop the necessary multi-source estimation methods.

The sample will then be redesigned to be drawn from the new SBR, taking account of the use of unit level VAT turnover data: a reduced sample size relating to larger or more complex statistical units only. On current plans we aim to have the redesign ready and the SBR MVP developed by March 2018 in order to facilitate a six month period of impact assessment analysis. This is needed to understand the impact on published retail turnover time series feeding into key short term economic indicators and Gross Domestic Product (GDP).

Business Index and Statistical Business Register

We are currently in the development phase.

- The business index is currently a beta live trial service, being used by six other UK government departments.
- The SBR is in its Alpha phase (due to complete at the end of September) then move into Beta build phase.
- Following the enactment of the Digital Economy Act 2017, negotiations on the delivery of new data sources from the UK Revenue department are progressing.

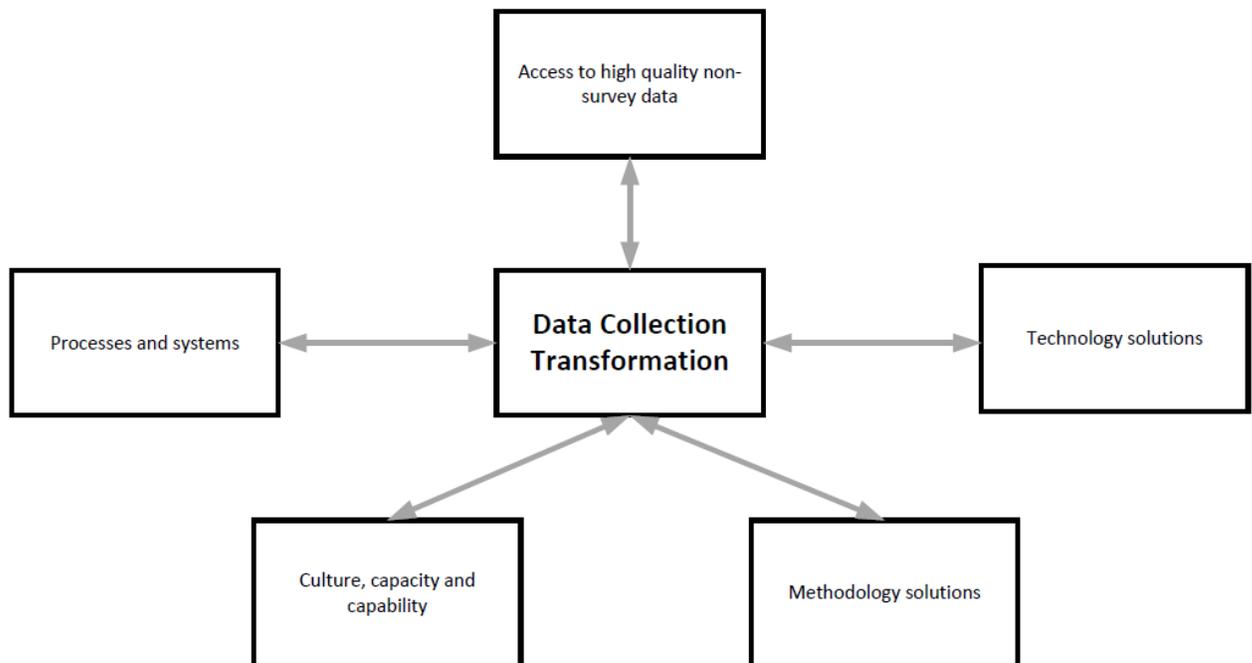
Direct Data Feeds from Businesses

Limited progress has been made to date due to a lack of funding. However recently progress has been made in moving towards agreement to conduct a pilot data collection on a monthly basis from large retailers. This will provide early benefits in terms of understanding statistical and technology requirements. It will also provide insight into data comparability collected via the different modes of online completion versus direct data transfer from businesses' own systems.

Dependencies across ONS

This approach of developing generic processing platforms for shared use across business surveys, social surveys, Census and the resulting statistical outputs presents huge opportunities for efficiency and standardisation, but also there are complex dependencies across the different areas of the organisation.

The following diagram summarises the key dependencies which business survey transformation is dependent on.



Questions for Discussion

- i. **The single platforms approach** generates important cash and efficiency savings; however it also creates significant cross-ONS dependencies and resourcing challenges. This creates the need for effective prioritization and regular review of development across the organization. ***Have other NSIs taken a similar approach to generic systems build across the organization? What have they learned from that experience?***
- ii. **The quality of administrative data and other non-survey sources** is still relatively unknown as ONS has not been able to access these datasets in the absence of appropriate UK legislation. Now that the new Digital Economy Act has been passed, access is expected to be attained from early in 2018. However extensive analysis remains to be done on how to replace surveys with these data in

collection and processing. *Many other NSIs have already implemented use of administrative (non-survey) data in the production of economic statistics outputs or National Accounts. How long did it take to complete the necessary research and development of methods to use a particular administrative data source with/instead of survey data for a given economic statistics output eg estimates of monthly or quarterly turnover?*

- iii.** **The success of obtaining data directly from businesses' own accounting or other electronic systems** given, for example differences in definitions or frequency, or due to technology challenges. *For NSIs that have implemented this, how long did it take to develop the data transfers and solve any statistical issues? Have NSIs got an example of how they overcame a difference in survey and directly transferred data for a key variable such as employment or turnover?*
- iv.** **The ability to maintain statistical utility of linked datasets**, and trialling new statistics needs sufficient time and capacity for robust testing. *What is other NSIs experience of linking data at unit level across datasets from different sources through the use of registers? For example linking economic and social variables?*
- v.** **Additional staff and skills are required** to satisfy the substantial demands placed on methodology and digital resource by ONS' multiple transformation programmes. *How have other NSIs managed the increase in expert resources in methods required to support transformation of business surveys? Have they recruited more people into the organisation or out sourced the work? If the latter, do they have an example of where this has worked successfully?*
- vi.** **The extent to which existing processing systems should continue to be enhanced**, as priority has been given to developing new strategic systems. In some cases, however, early transformation will require changes to legacy systems in order for the data flow to work effectively or new user needs arise that need to be met in the interim whilst new platforms are in development. *What has the approach been of other NSIs in this situation?*
- vii.** **Data security reduces statistical utility** and with increasing access to administrative data there is a greater than ever focus on measures required to hold and link data securely. By adopting these measures there is a risk that access to the required data will be limited, reducing the statistical utility of the data. *Where NSIs are linking multiple datasets for analysis by users, are there examples of where they have been successful, and unsuccessful, in being able to provide data at a low enough level of detail without compromising data security or disclosure of individual records?*
- viii.** **Maintaining high standards in statistical methods** is paramount when transforming quickly to make efficiency savings. *How have other NSIs balanced the need for developing sound new methods associated with using administrative data with the need to transform quickly?*
- ix.** **Obtaining buy-in from key external stakeholders** such as those organisations with responsibility for UK economic policy-making, to ensure that the impact of transformation on business data collection

is sufficiently discussed and accepted in the context of resulting discontinuities in key published economic time series. ***What sort of order of revisions have been made to key economic time series as a result of using administrative data sources? Do other NSIs have experience of where a lack of acceptance by external stakeholders has prevented the release of revised key economic time series due to moving from use of survey data to administrative data?***

Debra Prestwood
Head of Business Surveys Transformation Division
Office for National Statistics, United Kingdom
Email: debra.prestwood@ons.gov.uk
Tel: 0044 1633 455882