Transformation of UK Business Data Collection

UNECE Workshop on Statistical Data Collection
10-12 October 2017

Debra Prestwood
UK Office for National Statistics
Divisional Director of Business Surveys Transformation
Outline

• Vision for data collection
• Enterprise Architecture
• Roadmap for business surveys data collection
• Focus on short term surveys/statistics
• Statistical Business Register
• Direct data feeds
• Progress
• Sharing experience
Target Operating Model (TOM)

- **Data integration**, not simply data collection

- **Link core sources (survey and non-survey)** to a **central statistical spine** at a **micro data level** in order to:
  
  - **Maximise the utility of our data**
  
  - **Work efficiently** - link once, rather than multiple times in different locations for different purposes / outputs
Survey and non-survey integration

On-going survey redesign and rationalisation work is currently looking at how we can minimise the number of businesses we survey through the integration of various data sources.

Survey Data
Currently collected through 80 business surveys ranging from monthly, quarterly and annual.

Non-survey Data
Use the new Digital Economy Act (data sharing legislation) which allows ONS to store and use admin data, e.g. PAYE, VAT etc., plus commercial data.
Business Survey Redesign

Conceptual Approach

1. Harmonise concepts & variables

2. Modularise by association & periodicity

3. Rationalise by integrating other sources

4. Bespoke eQ modules for individual businesses depending on what data is already available from non-survey sources

NB. Module / survey names are for illustration purposes only
Vision

**Non-Survey Data**
- are preferred source
  - Value Added Tax
  - PAYE RTI
  - Self Assessment...

**Business Surveys** redesigned
to fill gaps or for calibration
- Harmonise
- Rationalise/Online modular design
- Sample redesign

**Statistical Business Register**
- Provides statistical units and so spine to link all/most business data
- Provides sampling frame for business survey(s) to supplement non-survey data sources
<table>
<thead>
<tr>
<th><strong>TRANSLATE</strong></th>
<th><strong>TRANSFORM</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>output(s) ‘as is’</strong></td>
<td><strong>revisit output requirements</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>TRANSLATION</strong></th>
<th><strong>RATIONALISATION</strong></th>
<th><strong>INTEGRATION</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>• No output changes or sample redesign</td>
<td>• Revisit output requirements</td>
<td>• Revisit output requirements</td>
</tr>
<tr>
<td>• Questionnaire content and structure largely unchanged</td>
<td>• Survey(s) redesigned and rationalised</td>
<td>• Register spine to integrate sources</td>
</tr>
<tr>
<td>• Improve questionnaire design by:</td>
<td>• Harmonisation of outputs, concepts</td>
<td>• Draw first on non-survey sources to meet output requirements</td>
</tr>
<tr>
<td>• Benefit from features of online collection (routing, validations, usability)</td>
<td>• Improved data collection design (content and usability – online first)</td>
<td>• Surveys redesigned to fill gaps</td>
</tr>
<tr>
<td>• Meets accessibility needs</td>
<td>• Amended processing systems and methods</td>
<td>• Online first data collection redesign</td>
</tr>
</tbody>
</table>

**Benefits / Value**

**TRANSLATE**
- Reduced costs - GAE
- Reduced burden (but limited to usability)
- Reputation – ‘digital by default’

**RATIONALISATION**
- Revisit output requirements
- Survey(s) redesigned and rationalised
- Harmonisation of outputs, concepts
- Improved data collection design (content and usability – online first)
- Amended processing systems and methods

**INTEGRATION**
- Revisit output requirements
- Register spine to integrate sources
- Draw first on non-survey sources to meet output requirements
- Surveys redesigned to fill gaps

**Benefits / Value**

**RATIONALISATION**
- Reduced costs - GAE & staff
- Reduced burden - sample reduction, improved questions, usability
- Reputation – progressing to transformation vision, ‘digital by default’

**INTEGRATION**
- Reduced costs - replacing direct collection, GAE & staff
- Reduced burden - data re-use, sample and survey reduction, improved questions, usability
- Reputation – delivering transformation
- Promotes innovation, new outputs
- More responsive – flexibility to respond to new demands
# Potential impact of new sources on existing ONS business statistical outputs

<table>
<thead>
<tr>
<th>New Source</th>
<th>Potential impact on ONS outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAT</td>
<td>Redevelopment of smaller MBS Quality assurance of many business surveys VAT expenditure=intermediate consumption proxy</td>
</tr>
<tr>
<td>HMRC Trade &amp; VAT</td>
<td>Improve samples for ABS, environmental accounts, prices, Prodcon</td>
</tr>
<tr>
<td>Corporation Tax</td>
<td>Redevelopment &amp; quality assurance of ABS, R &amp; D, Capital (GFCF)</td>
</tr>
<tr>
<td>Self-Assessment</td>
<td>Enriched Statistical Business Register Improved current measure of self-employment</td>
</tr>
<tr>
<td>PAYE RTI</td>
<td>Redevelopment of workforce jobs Redevelop MWSS and ASHE</td>
</tr>
<tr>
<td>Visa, other credit cards</td>
<td>Add trade mode of supply Improve regional accounts</td>
</tr>
<tr>
<td>Scanner data</td>
<td>Redevelopment of RSI &amp; business prices</td>
</tr>
<tr>
<td>Solvency II</td>
<td>Redevelop insurance surveys</td>
</tr>
</tbody>
</table>
Potential timeline / road map
Emerging road map for transforming all business survey / data groups
Why start with short-term stats?

Non-Survey Data are preferred source
- VAT
- PAYE RTI

Business Surveys redesigned to fill gaps or for calibration
- Online modular design
- Harmonise

Statistical Business Register
- Provides statistical units and so spine to link all/most business data!
- Provides sampling frame for business survey(s) to supplement non-survey data sources!
- Turnover (VAT) and employment (PAYE) are core register (auxillary) variables!
...so great for laying the foundations for transformation!

• High potential for use of non-survey data (more so than structurals)
• We already have VAT data and are researching PAYE RTI potential
• VAT and PAYE RTI are the initial focus for non-survey data integration
• Makes use of investment in the ongoing VAT Minimum Viable Product
• High cost surveys due to size, coverage, frequency
• Short frequency to reiterate and learn

• High potential to harmonise turnover (and emp)
• Relatively simple concepts and questions compared to structural surveys
• RSI, QBS, MWSS are already online, so functionality in place for ‘simple’ surveys
• All use same downstream legacy systems

• High potential to use of non-survey data (more so than structurals)
• We already have VAT data and are researching PAYE RTI potential
• VAT and PAYE RTI are the initial focus for non-survey data integration
• Makes use of investment in the ongoing VAT Minimum Viable Product
Workstreams

Output Specification
Define & prioritise **output user needs**, to provide a reference for all other workstreams

**Non-Survey Data**
- are preferred source
- VAT
- PAYE RTI
- Self Assessment...

**Business Surveys** redesigned to fill gaps or for calibration
- Harmonise
- Rationalise/Online modular design
- Sample redesign

**Statistical Business Register**
Provides statistical units and so spine to link all/most business data
Provides sampling frame for business survey(s) to supplement non-survey data sources
Workstreams

Output Specification
Define & prioritise **output user needs**, to provide a reference for all other workstreams

For example...
Who are our main users for short-term output (turnover) data? What conceptual basis should we be using to collect data and/or produce stats?

Global National Accounts vs. BoE– Monetary Policy Committee

**Economic Ownership**
- UN System of National Accounts
- Who owns the raw materials & finished goods?
- Where is the money going?

What is the ‘real’ activity in the UK?
Use of ‘Simple v Complex’ Units

• Motivated by desire for admin (or other non-survey) data to be the preferred source [for any given output]

• Suitability depends on quality (inc. availability) of admin data for the statistical units

• Likely to want to partition business population into units for which:
  • non-survey data will be used
  • survey principles will be applied
Use of Simple v Complex Units

Size of unit

Big

Small

Complexity

Simple

Complex

defined primarily by:
unit linkage
(1-to-1, 1-to-many, etc.)
defined by:
employment, turnover, …
New Statistical Business Register

**Business Index (BI):** Links legal units to VAT references and PAYE schemes

**Statistical Business Register (SBR):**

- **Frame function:** Provide the sampling frame facilitating survey and non-survey sources. Enable more sophisticated and flexible sample designs and auxiliary variables for estimation
- **Statistical spine:** Create statistical units and so the foundations and building block on which to integrate data at microdata level and produce statistics
- **Analytical function:** Improve analysis of business population through holding more variables and most up to date data
- **Expanded population:** Include Self Assessment Tax ie very small sole traders and partnerships
‘Direct data feeds’ from businesses

Research into:

• Compatability with company accounting/processing systems
• Accessibility
• Software barriers
• Flow alongside survey data into common platform(s)
• Other NSIs’ experience
Progress

- 5 business surveys ‘translated’ online
- Equates to 68,000 eQuestionnaires to businesses
- 12 surveys using new Secure File Transfer portal
- Next stage is translation of rest Monthly Business Survey
- Short term statistics transformation in progress
- Business Index in live beta trial phase
- SBR in design Alpha phase

Monthly Business Survey
We’d like to understand your experience in…

- Generic platform build across multiple statistical domains (Business, social, Census)
- Length of research time into use of admin data source to replace survey data
- Obtaining ‘direct data feeds’ from businesses
- Linking economic and social microdata through registers
- Maintaining security of data when linked at microdata level
- Capacity to manage transformational methods research – outsourcing to third parties used?
- Moratorium on change to existing systems and processes whilst transforming data collection?
- Ensuring sound methods for integrating admin sources v transforming quickly
- Handling of resulting discontinuities in key economic time series
Thank you for listening!

Contacts:

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