

Meeting of the Group of Experts on Business Registers
27-29 September 2017
OECD, Paris

The role of the new SBR in the UK statistical production process.

Andrew Allen, Office for National Statistics, UK

Introduction

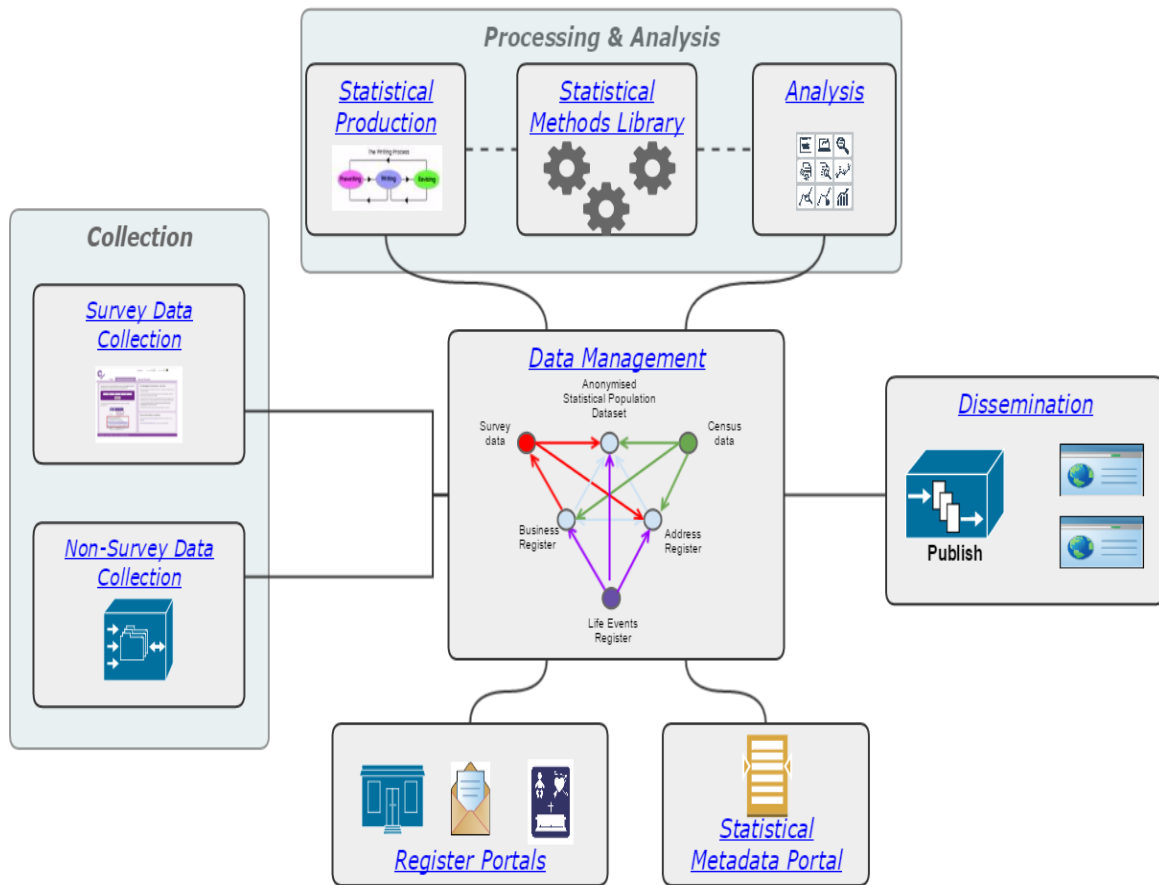
A major transformation programme ranging from data collection to economic statistics production is underway at the ONS. As part of this programme, a new statistical business register is being developed, as well as a new legal unit based list of the business population, called the business index. The new Business Index and Statistical Business Register will replace the current Inter-Departmental Business Register.

The current business register is based on 20 year old technology, and is a standalone database. The new SBR is a key to the ONS IT strategy; it will be created within our Cloudera technology environment and will become a fundamental component of the ONS enterprise architecture. This environment will include a data management platform and a number of services used to process data. The new SBR will be built using these platforms/services, and hence will not be a standalone register database. A new user interface will enable the normal register functions to be carried out using these platforms and services.

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 wider use a new product called the Business Index is currently being trialled (in private beta). 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, but also serves as the first step in the business register process. The new SBR will also provide a spine function enabling linking of business datasets and facilitating the use of administrative and survey data, to create statistical outputs.

New Technology Environment

ONS is transforming its information technology landscape to a service oriented architecture, with Cloud Foundry hosting its services and Cloudera providing a data management and processing platform. The old IDBR is being replaced, so rather than building it in an isolated database, we will be using this new infrastructure. The consequence of this is that the new register will be based on different platforms and built up from a range of services. To the end user it will look like a single harmonised system, but in reality it will be a network of systems.



The diagram above summarises the enterprise architecture being introduced to ONS. The core products from an SBR perspective are:

Data Management platform - This is where all of the administrative data to be used in the formation of business entities for the SBR will be ingested and held. Data created specifically for the SBR, such as local units would also be stored in this platform.

Statistical Production platform - This will be where the execution of SBR methods are orchestrated, i.e. linked together to provide the necessary service.

Statistical Methods library – will store the methods coding needed for SBR and other statistical production.

Meta data - The new approach relies heavily on Meta data. The data management platform will create this Meta data and also the versioning of, and secure access to all our data.

The new IT environment allows for a large improvement in data handling efficiency, for example data can be stored once and used many times, without having to physically move it around the organisation. It will also mean that in future we will not be dependent on a single large difficult to maintain system. The SBR will effectively become modularised, making future updating and changing easier to manage.

Expanding the Role

Business Index

The IDBR created enterprises directly from administrative units, without firstly creating legal units. The new system will first create a legal unit view of the business. This is called the BI, and will operate as a separate service, providing a cross government legal unit identification service. The BI will then feed these legal units into the new SBR system, where the family of statistical units will be created.

This expands the role of the register, because users will be able to see the legal view of a business, which in many cases is more relevant than the economic view provided by the enterprise view. Ideally this service could be used for a wide range of operational services, and our research and user engagement has identified a large demand in the UK for this type of service. However, there is currently no legal gateway to enable the service to be used for non statistical purposes. Even so, a number of other government departments have started to use the BI service, because they want to have a legal view of the business community. Example uses include deriving management information about an operational service, by appending BI variables to management information.

SBR

The role of the SBR will change within the modernised ONS. The IDBR functioned mostly as a sample frame, but the greater use of administrative data in future will enhance the role of the SBR. The future system will need to provide a frame which facilitates mixed data acquisitions, including both traditional 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 that 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.

As part of the development, the scope of the SBR has been re- considered. The main point of discussion is around whether the new SBR should also include the sampling functionality (which was included in IDBR). Taking an enterprise architecture approach indicates that the sampling function goes beyond the scope of the SBR. However , since the relationship between frame and sampling is close , the SBR project team will build the necessary sampling functionality. This will be built as a separate process to follow the service based architecture.

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 , particularly to other government users.

Analytical function

For many years there has been a growing demand to use the business register for economic analysis of the business population. The IDBR had a number of shortcomings due to its original design being

primarily as a sampling frame. These mostly concerned the under coverage of units , age and meta data for auxiliary variables. The new SBR is being designed with the analytical communities needs in mind. The most important of these are very clear Meta data and time- stamping for each variable. Secondly we aim 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.

For variables such as employment, turnover and classification, different versions could be held to satisfy the different user requirements. For example, the latest monthly employment from the administrative system (PAYE) would be of most use to the analytical community , but for survey stratification , a more stable annual average of employment might be more suitable .

Holding the most up to date version of variables will pave the way to use the new SBR to provide direct statistics i.e. potentially to be able to produce register based statistics, but this will be explored further in future.

Expanded population

The IDBR was known to have under-coverage of very small entities. This was managed through estimation using aggregate data from administrative sources, but the wider role requires this under-coverage to be addressed. ONS has also recently benefitted from new data sharing provisions (Digital Economy Act 2017), 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, whose sales are below the £83,000 VAT threshold and who do not employ any other staff. 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. However, there is a growing policy requirement to analyse this population in detail. The expanded population will allow detailed industrial and geographical analyses of this part of the population for the first time.

Spine function

The increased use of administrative data also creates a business spine function for the SBR. By enabling different data sets to be linked via the spine new statistics can be produced, as well as new methods to produce existing 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, providing better industrial classification, enabling the production of outputs.

The new IT infrastructure will facilitate this role, by storing data in a central repository , making linking and reuse more practical. Using the most update to data, backed by very good time-stamping and Meta data are also important parts of the new design , to facilitate this use.

Progress to date.

We are currently in the development phase. Progress to date includes:

- The business index is currently a beta live trial service, being used by six government departments.
- The SBR has just finished its Alpha phase and is seeking to move into the Beta phase. The first deadline will be to deliver a minimum viable product for ONS short term survey redevelopment in spring 2018.
- Following the enactment of the Digital Economy Act, negotiations on the delivery of new data sources from the Revenue department are progressing.