Wired for data – the transformation of data collection in the UK.

Kevin J. Moore

Office for National Statistics, United Kingdom, kevin.moore@ons.gsi.gov.uk

1. Introduction

The UK Office for National Statistics (ONS), along with the rest of the UK Civil Service is coming under increasing pressure to become more efficient. At the same time, the thirst for statistics is growing. In response, the ONS is investing in a multi-pronged transformation of its data collection capacity.

This paper sets out the vision and aims of the transformation. It describes how one branch within the Business Data Division (BDD) of the ONS, has implemented a generic system that goes some way to fulfilling the needs of the programme. It also highlights some of the challenges in co-ordinating the programme that have come out of the lessons learnt during the implementation of the new system.

2. Background

The BDD is responsible for the collection and validation of unit level data, from businesses across the UK. It deals with approximately 80 annual, quarterly and monthly business surveys and over 1.7 million survey questionnaires are despatched to approximately 320,000 businesses each year.

The business survey data collected by BDD feeds into the production of the key economic indicators produced by ONS, used for economic policy making by the UK government. They are also used by a wide range of users, including academics, industry and the public. The data collected by BDD also enables ONS to meet a range of EU regulations.

One such EU regulation is the European System of Accounts. This regulation affects the collection from ONS Financial Inquiries (FI). The 2010 revision to the accounts (ESA10) required a substantial redesign of the FI and an updated system to process them. The

decision was made to redevelop the Common OpenRoad Architecture (CORA) system to facilitate the redesign.

The BDD ran a project to redevelop CORA using Agile project management techniques. It brought together a project team that included IT professionals and representatives of the eventual users of the system. Development of functionality for the platform was prioritised based on the benefit it provided to the platform.

The Data Collection Transformation Programme (DCTP) in ONS is changing the way ONS collects and uses data to ensure the most efficient statistical production process. The scope of the programme includes systems and process to make use of survey and Administrative data collected digitally by default.

3. The Challenge

The goal of a redeveloped CORA platform was to standardise business processes, rationalise statistical output production methods, reduce risks associated with systems and enable innovation. This goal, however, had to be met alongside the need for producing the ESA compliant statistical outputs.

The vision of CORA to achieve its goal was to develop generic functions that are configurable by the user which can be applied to any survey. In doing so, the functions would only need to be developed once. They would have a specific use and linked together to enable the steps required to fulfil the specific steps in the Generic Statistical Process Model (GSBPM).

The customers for CORA are all internal to ONS. The final system would enable the collation and validation of micro data from multiple data sources by BDD. That data is then delivered to our Results, Analysis and Publication colleagues to an agreed timetable and at a quality level appropriate to their needs and within budgetary constraints.

With fixed time (ESA regulation timetable) and budgetary constraints, it had to be scope that could flex. It is the move to generic tools and functions that provided that flexibility.

4. The CORA principles

In assessing the suitability or the effort to get a survey on to the CORA platform, a number of key questions are asked. These questions are intended to identify, at a high level, the gaps in functionality of CORA compared to the current survey process. These are obtained via a proforma which is then discussed with the customer. However, during these discussions, the following principles are adhered to:

- Is the functionality required?

Survey processes evolve over time to resolve particular nuances in the survey. As part of the rationalisation of methods, any process that has not already been addressed is challenged. Scenarios where the processes are used are examined to ensure the resolution is statistically sound (increasing data quality).

- Have these scenarios been addressed already?

During the assessment of the requirements and scenarios provided by the survey representatives, it is possible that a similar scenario has already been addressed as part of development of an earlier survey. Instead of developing something completely new, can functionality already on the CORA platform be reused or enhanced to meet the needs of the survey? Is there a trusted process already in ONS that resolves these scenarios and can that be integrated into the platform?

5. Generating Solutions

As indicated, the CORA project team consists of representatives from across the survey process including IT and Methodological professionals. Agile principles ensure iterative refinement of requirements which in turn affect the proposed solution. A high amount of inclusion is sought from all members of the team but at different times. This requires an effective communication structure framework.

Initially, a workshop is held with high ranking members from the IT professionals, the CORA product owner and a business analyst. This is to understand the requirement in general terms and raise any uncertainties over the need for the function as per the CORA principles outlined in section 4. If required there will be an agreement between them over possible solutions including the reuse of available functionality if appropriate.

The proposed solution is discussed with the survey area that has the requirement. This may also include a demonstration of a prototype solution. This stage irons out any questions or uncertainties over the solution or behaviours of the system under different conditions. It is at this stage where acceptance criteria are drafted to facilitate sign-off of the solution as complete.

The agreed solution is described to the rest of the IT team in the planning sessions for each sprint. These contain the creation of "stories"; deliverable packages of work. The progress of these packages is then monitored through the development cycle.

Following completion, a demonstration of the solution is performed to the customers of all the stories in the sprint. The solution is tested against the acceptance criteria of each story to ensure the requirement of the customer has been met. Customer acceptance testing is undertaken by all surveys on the platform before the new functionality is put into use in the live environment.

Following a sprint, a retrospective is held to identify any changes required to the process to help improve them in future. Metrics of project activity are recorded and fed through the project governance to ensure project direction. An example one metric monitored, the sprint burn up, is illustrated in figure 1.

Figure 1 – Example of a sprint burn up chart.
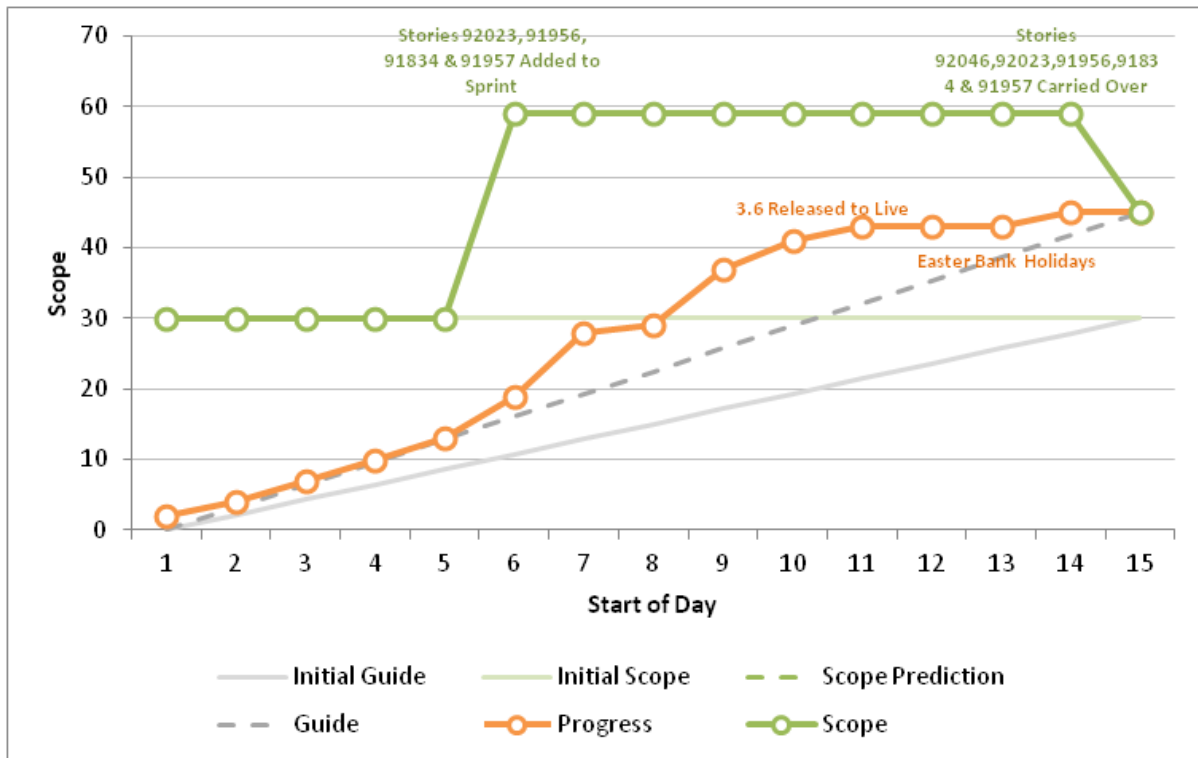
6. Quality, quality, quality

The Senior Responsible Owner for the CORA platform is also the Champion for Quality in ONS and the Deputy Director of BDD. Therefore producing a quality output has been at the forefront during the CORA project. Despite the co-ordination of processes and the other goals of CORA described in section 3, it is imperative CORA provides the tools to enable BDD to continue to meet quality targets in the Service Level Agreements with their customers.

A suite of validation rules has been developed to facilitate data cleaning. They consist of structural and response validation rules. Structural validation ensure data received by ONS is what was expected with regards to the shape of the data, response rules are performed at micro data level and test for "strange" responses; higher or lower than expected, extreme movement from previous period/year etc.

Surveys using the CORA platform are collected into families with similar characteristics. Each family is driven by questionnaire and question libraries that are reusable across all the surveys within the family. Both questionnaires and questions are time bound within a survey and enable initial structural validation of responses to the survey during data take on.

Further structural validation is performed on the respondent themselves. An automated load of the sampled businesses is performed for each survey period from the sampling frame and matched to those returning data.

Questions within the library for a survey contain attributes that are maintained by the survey such as; is unique, is essential, is repeating. These attributes are used within the validation rules associated with that question.

7. Achievements

CORA has enabled ONS to meet its European commitments as part of ESA10. The compliant Financial Services Surveys were the first to use the new system. Following that, ONS has reintroduced its Purchases Survey after over a decade and uses CORA to clean the data.

CORA was the first project in ONS to use Agile project management. The success of the project and the implementation of Agile was, in no small part, due to the collaborative team built up across all areas of the office. The CORA blueprint for working in this way is now being adapted for use in other projects and programmes including DCTP.

## 8. The enhancement of CORA into DCTP

As successful as CORA has been in delivering a data take and validation platform, ONS needs to transform its data collection capability further. In general, ONS surveys are currently performed by paper questionnaire. There are numerous processing systems depending on the survey; some of which date back to the 1990s. In order to become a modern, efficient Government department substantial change has to take place.

DCTP aims to change business processes, make effective use of non-survey data in ONS statistical outputs, and collect survey data online. It plans to achieve these by developing an integrated data store; linked across data sources and common IT services and systems built with the latest tools and methods. At the same time reducing the burden on respondents.

What has been achieved by CORA and the lessons learnt from it will feed into the "common IT services and systems" stream of DCTP. The vision is that the new ONS digital services

will be organised in 8 – 12 platforms made up of individual products/services. A diagram representing how the new platforms will link together can be found in figure 2, below.

Figure 2: Representation of DCTP platforms.
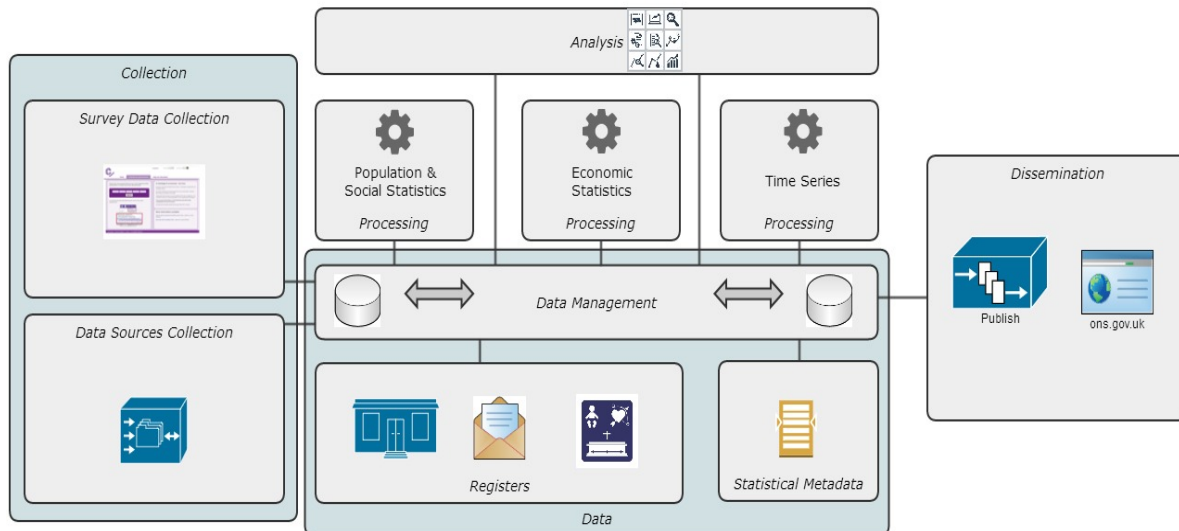
The function being performed by CORA maps directly to the requirement for a data validation service as part of the Survey Data Collection platform. However, due to the generic templated nature of the CORA solution to validation, there are also synergies with the validation of non-survey data. The role of the CORA team going forward will be to ensure what has been learned in developing the CORA validation suite is fed into the DCTP plans for that service.

9. Conclusions

- What is trying to be achieved by the transformation of data collection in ONS is a mighty challenge. However, what has been already delivered by the CORA project is an indicator of the efficiencies possible.
- It is possible to implement generic and reusable processes and procedures in statistical output production independent of the nuances associated with the surveys themselves.
- The success of CORA hinged on the use of Agile working and the inclusion of the users of the system during its development. It is not unreasonable to assume the Data Collection Transformation Programme in ONS will be any different.