Contents of Journal of Official Statistics: March, Vol.29, No.1, 2013

Journal of Official Statistics, Vol. 29, No. 1, 2013, pp. 5–28

Consolidation and Standardization of Survey Operations at a Decentralized Federal Statistical Agency

Jack Nealon and Elvera Gleaton[1]

With tighter federal budgets on the horizon, the National Agricultural Statistics Service decided in 2009 to pursue three architectural transformations, primarily to provide savings in staff resource costs by enabling the centralization or regionalization of survey operations. The transformational initiatives involved: (1) centralizing and consolidating network services from 48 locations; (2) standardizing survey metadata and integrating survey data into easily accessible databases across all surveys; and (3) consolidating and generalizing survey applications for the agency's diverse survey program. The three architectural transformations will be described as well as initial efforts to consolidate and standardize survey operations across the agency.

Key words: Virtualization; survey metadata; analytical database; transactional database; generalized application services.

1. Introduction

The mission of the National Agricultural Statistics Service (NASS) in the U.S. Department of Agriculture (USDA) is to provide timely, accurate, and useful statistics in service to U.S. agriculture. Each year, NASS conducts hundreds of surveys and produces over 400 agricultural statistics reports, which are disseminated on a weekly, monthly, quarterly, or annual basis and cover most aspects of U.S. agriculture. NASS also performs reimbursable agricultural surveys, primarily for other USDA agencies, which make up approximately ten percent of the NASS budget. Finally, NASS is responsible for the Census of Agriculture every five years.

NASS has managed its diverse survey program through a network of 46 Field Offices and a Headquarters Office in Washington, D.C. and Fairfax, Virginia. A Field Office is located in Puerto Rico and in every state, except five New England states that are serviced by a single New England Field Office in New Hampshire. The Field Offices work closely with the State Departments of Agriculture and land-grant universities[2] through Federal/State cooperative agreements to collect additional agricultural data important to local economies, such as county data. The Federal/State cooperative agreements

[1] National Agricultural Statistics Service, United States Department of Agriculture (USDA). USDA-NASS, 1400 Independence Ave., SW, Washington, DC 20250, U.S.A. Email: Elvera.Gleaton@nass.usda.gov

[2] Starting in 1862, the U.S. Federal Government granted federally-owned land (hence the name "land-grant") to states for the development of land-grant universities offering education in agriculture, engineering, and traditional academic subjects.

minimize duplicate data collection efforts between Federal and State agencies, lighten the burden on survey respondents, and foster consistency in the production of agricultural statistics.

Through its 46 Field Offices, NASS uses a local presence to build trust with data providers and credibility with data users to encourage support and cooperation for NASS surveys. The 46 Field Offices help increase awareness of the NASS survey program throughout the country and keep the NASS Headquarters Office informed about local agricultural issues that can potentially impact surveys. On the other hand, the NASS decentralized structure makes it an ongoing challenge to ensure survey procedures are performed consistently across the 46 Field Offices. Field Offices vary in size – from only one federal employee in Alaska to 32 federal employees in California. Most Field Offices have historically employed statisticians and statistical assistants to execute several survey functions for that Field Office, such as list sampling frame maintenance, data collection, and data review.

Data are collected from farmers and ranchers, livestock feeders, grain elevator operators, and other agribusinesses, who voluntarily supply the data in a variety of ways – mail questionnaire returns, web data collection, telephone interviews, face-to-face interviews, and crop measurements. The National Association of State Departments of Agriculture (NASDA) has traditionally provided about 3,000 part-time field and telephone interviewers to collect the survey data for Field Offices. These interviewers, especially field interviewers, often have knowledge of the local agriculture and are known in the community, which can be an asset when interviewing sampled farm operators repeatedly.

The global economic crisis that surfaced during 2008 led to discussions that many federal budgets would become tighter in the future to address the expanding federal budget deficit. NASS realized it must operate more efficiently to remain relevant as a federal statistical agency under the projected tighter budgets. At the beginning of 2009, NASS Administrator Cynthia Clark made a formal request to all employees asking for suggestions on ways to be more efficient as an agency. Two main themes were identified from the input that could make NASS not only more efficient, but also potentially more effective as an agency: improving the survey processing architecture and the staff resource utilization.

1.1. Survey Processing Architecture

An inefficient decentralized survey processing architecture had evolved across the 46 Field Offices for several survey functions. For example, each of the 46 Field Offices had to make updates on their Local Area Network (LAN) whenever a change was made to a Blaise[3] data collection or data editing instrument or to a SAS[3] data analysis application, so the task was performed 46 times rather than potentially only one time. Thousands of Blaise and SAS data files were created, manipulated, and transferred from application to application each year when it would have been more efficient to store the data in an easily accessible, centralized database. The decentralized survey processing architecture led to

[3] Blaise is survey software from Statistics Netherlands (www.blaise.com) and SAS is analytical software from SAS® (www.sas.com).

the underutilization of a wealth of survey data located in thousands of disparate and proprietary data files since analysis could not be readily performed across years, surveys, and commodities. In addition, the decentralized survey processing architecture limited the implementation of new survey procedures. As an example, NASS had a goal of developing customized survey questionnaires for farmers and ranchers to reduce respondent burden, but this could not be realized due to the poor integration of survey data and metadata that had evolved from the decentralized processing architecture.

1.2. Staff Resource Utilization

NASS is unique among federal statistical agencies in the U.S. since it has a decentralized structure for conducting surveys using 46 Field Offices. However, the decentralized structure contributed significantly to almost two-thirds of the total NASS budget in 2009, being devoted to salaries and benefits for employees. Could NASS operate more efficiently by performing survey functions at a central or regional office rather than separately in 46 Field Offices? As an example, list sampling frame maintenance of the farm population has historically been performed by one or more employees in each Field Office for that Field Office. Could many of the list sampling frame maintenance tasks be performed centrally using fewer statisticians and statistical assistants than used in the Field Offices? In addition to providing human resource savings, centralization or regionalization of various survey functions from the 46 Field Offices would also provide the opportunity for more consistent or standard survey procedures, which potentially could improve the quality of agricultural surveys.

Unfortunately, NASS was not able to explore ways to perform survey functions in a central or regional office structure, since the decentralized technology architecture would not effectively support it. Therefore, NASS identified the following three foundational and architectural transformations to be addressed with high priority from 2010 through 2013:

(1) ***Network:*** Centralize and consolidate network services across the agency to remove location as a barrier to shifting or combining work across Field Offices.
(2) ***Metadata and Data:*** Standardize survey metadata and integrate survey data in easily accessible databases to streamline survey operations from design through dissemination.
(3) ***Applications:*** Consolidate and generalize survey applications and make optimal use of the enterprise databases so surveys are processed in a standard manner and survey functions can be performed effectively from any location in the agency.

These three architectural transformations, once implemented, would finally position NASS not only to use staff resources more efficiently by *centralizing or regionalizing critical survey operations*, but also to potentially perform survey functions more effectively through *more standardized survey operations*.

In the next three sections, the objectives and benefits, transformational architectures, and lessons learned will be discussed for the three initiatives. Then, the initial consolidation and standardization of survey operations at NASS, which were enabled by these three initiatives, will be described in Section 5.

2. Objectives and Benefits

The overarching objective of the three architectural transformations is to provide NASS with the ability to centralize and/or regionalize survey operations, primarily to reduce staff resource expenses. However, there are additional objectives that will now be described for each of the three transformations that will provide significant benefits to NASS. NASS used the following four criteria to determine if each transformation eventually met its objectives: the transformation enabled *centralization, integration, standardization, and flexibility*. By satisfying these four criteria for the three transformational initiatives, NASS believes it will be able to continue to achieve the agency's mission during times of tighter budgets by utilizing staff resources more effectively, potentially reducing survey errors, shortening the time period to complete many surveys, and providing improved data products to the public.

2.1. Network Centralization and Consolidation

NASS realized it had to migrate first from its distributed local area network (LAN) environment to a centralized environment to remove location as a barrier to conducting surveys. Therefore, the objective was to consolidate the LANs from 48 locations (46 Field Offices, Washington, D.C., and Fairfax, Virginia) to only two sites (production and disaster recovery sites). The goal of the consolidation of the network services was to make the agency more efficient due to the centralized, integrated, standardized, and flexible network design.

A *centralized* design means LAN administration would be performed by fewer employees from a central location, enabling NASS to reduce the staff resources devoted to LAN services. Computer and physical security can also be strengthened by greatly reducing the number of locations containing physical servers and electronic personally identifiable information. An *integrated* design allows all employees access to the same desktop image, making it more efficient to manage desktops centrally. A *standardized* design provides the same desktop image to all employees with the same role. As part of standardizing the server configuration, NASS also standardized directory structures, roles, and access rights, which is a very important element in the failover and disaster recovery capability the system provides. A *flexible* design means the new network architecture allows employees to access applications and data from anywhere at any time using a web browser from their desktops, laptops, iPads, or smartphones. This ease of access facilitates remote access to the network by employees.

2.2. Standardize Survey Metadata and Integrate Survey Data

Even after the network has been centralized and consolidated, the survey metadata would still not be consistent and the survey data would still not be well integrated. Therefore, NASS committed to migrating to consistent survey metadata and easily accessible, centralized databases. There were two main reasons: First, NASS wanted to minimize work inefficiencies caused by the decentralized data management process. Second, NASS wanted to provide employees easy access to integrated historical and current survey data to potentially improve survey operations by enhancing sampling and estimation procedures, simplifying survey management activities, improving edit and imputation processes, broadening analytical capability, and expanding dissemination products for data users. There are numerous benefits for NASS in implementing this centralized, integrated, standardized, and flexible database design.

A *centralized* design eliminates the need to create, manipulate, and transfer thousands of data files each year, such as Blaise and SAS data files, when the data resides in a centralized database. It also allows the Agency to operate as one rather than 46 separate Field Offices since survey tasks requiring access to survey data can be readily shifted from office to office, which makes better use of available staff resources. An *integrated* design provides standard metadata and enterprise databases that are shared across survey applications rather than applications having different proprietary data sources and metadata, which have introduced work inefficiencies and data errors at NASS. As an example, over 5,000 data errors were identified and resolved in NASS' published agricultural statistics data series when NASS transitioned to an integrated dissemination database in 2010. A *standardized* design, such as the dimensional design that will be described later for NASS' analytical databases, provides employees with easy access to many years of survey data responses, which expands their ability to conduct thorough data review and analysis. A *flexible* design allows for information to be stored in a single database, such as the published agricultural statistics from all surveys and censuses, which positions NASS to provide more diverse data products to the public.

2.3. Database Optimized, Generalized Survey Applications

During the past two decades, numerous survey applications were developed across the decentralized NASS organization that resulted in less than efficient survey processing. As an example, NASS had multiple applications doing the same types of edit functionality. In addition, hundreds of customized applications were developed in the Field Offices for State reimbursable surveys, which not only were an inefficient use of staff resources, but also introduced inconsistencies in survey processing. A more efficient survey process can result when there are fewer applications to maintain and fewer applications for employees to master. Also, an application needs to interface with a database design that is optimized for the survey function being performed by the application, such as optimizing the database design for timely retrieval of data for data analysis. NASS uses the term "database optimized, generalized survey applications" to emphasize the critical need that enterprise databases be designed to optimally service the needs of the generalized applications.

NASS committed to using database optimized, generalized application services for all surveys with the following three objectives in mind. First, re-engineer or enhance survey applications to make optimal use of the enterprise databases. Second, standardize about 150 nonstandard surveys across the Field Offices and use generalized applications for these surveys. Third, eliminate duplicative applications used in the Agency. NASS planned to modernize its survey applications through a centralized, integrated, standardized, and flexible application design.

A *centralized* design means generalized survey applications will be developed and maintained centrally for all Federal and State surveys. An *integrated* design ensures the same metadata and enterprise databases are shared across applications rather than applications having different data sources and metadata, as they have in the past. A *standardized* design provides standard application services across the 46 Field Offices for the 150 smaller NASS surveys that do not have a consistent survey processing environment. A *flexible* design allows employees to access the generalized applications through a thin-client interface. It will also simplify the process of updating or redesigning specific generalized application services at a later time.

In closing, NASS considered the many benefits from the three transformational initiatives as critical to the agency's success in the future. Therefore, NASS was willing to invest the funds over multiple years to pursue these challenging initiatives in order to potentially reap the many benefits as a federal statistical agency.

3. Transformational Architectures

The architectures for the three initiatives will now be discussed. These architectures were transformational since they significantly impacted upon the entire survey process and positioned NASS to explore more consolidated and standardized survey operations.

3.1. Network Centralization and Consolidation

NASS used virtualization technology to centralize and consolidate its decentralized network environment. Traditionally, a physical machine was used for each server deployed. With virtualization, a server or desktop functions as a software image, which means multiple virtual servers and desktops operate on a single physical machine (VMWARE 2006). The virtual desktop is simply the user's interface to a desktop that is stored on a remote server rather than physically on their desk.

The LAN centralization at NASS consolidated 94 physical servers across the country into 44 virtual servers in two locations (Gleaton 2011). Prior to the consolidation effort, Field Office and Headquarters LAN Administrators could load different software on a desktop machine depending on a user's request. This resulted in over 100 desktop configurations at NASS. The consolidation converted the agency's physical desktop configurations from over 100 to three virtual desktop configurations managed centrally in Headquarters. Figure 1 depicts the LAN centralization and consolidation process.

Virtualization is rapidly transforming the information technology work environment, since it allows agencies to run multiple virtual machines on a single physical machine. Through virtualization, employees have the ability to access applications and data from all 46 Field Offices rather than being limited to only using applications and data from their particular Field Office. Therefore, virtualization provides NASS with the ability to regionalize or centralize survey functions. NASS decided to virtualize the entire desktop rather than the more common technique of virtualizing individual applications to execute from a server. This decision was made as NASS determined that virtualizing each individual application was complex and inefficient due to hundreds of smaller legacy applications in the 46 Field Offices.

NASS engaged in considerable research and technical consultation to design and configure the centralized virtual environment. Detailed design, configuration, and implementation plans were prepared. The centralized solution consisted of two server

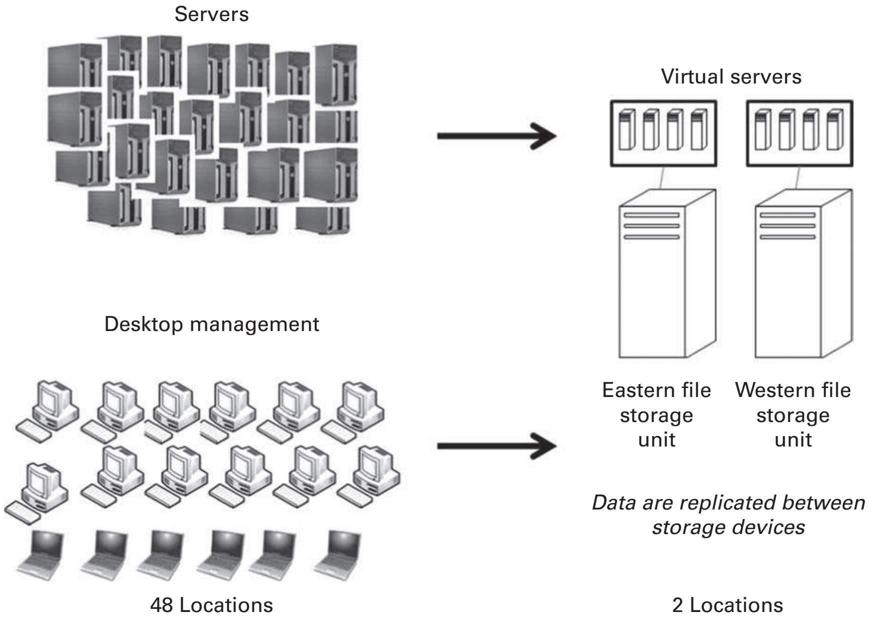

Fig. 1. Centralized and Consolidated LAN Services

farms in order to provide disaster recovery capability by continually replicating the data between the two sites. A server farm is a cluster of servers located in one site. The 46 Field Office locations were separated into two groups, with each group connected to either the eastern or western server farm. Employees from Headquarters and the eastern Field Offices were connected to the eastern server farm. A virtual desktop is sent to an employee's web browser from the server farm. The employee then connects to the centralized environment using a web browser. Generally speaking, no data are sent over the connection since only mouse clicks and keystrokes are sent over the network.

The virtualized environment consists of six critical components: *software, desktops, servers, telecommunications, storage, and disaster recovery*. The virtualization software enabled NASS to deploy virtual desktops to all employees. Virtual desktops look exactly like a physical desktop, are generated from one main image, and are distributed to all NASS employees. The virtual desktops are stored on servers in the two server farms. Employees are automatically routed to their virtual desktops when they log into the network and access NASS applications through a web browser. All software and applications are deployed through the virtualized desktops. The virtual desktops are read-only desktops so employees cannot install any software or devices. With the virtualization effort, NASS has started transitioning to thin-client desktop machines. These machines are inexpensive and usually require limited service while still providing full functionality because applications and processing are delivered from the servers rather than the desktops.

As noted, NASS uses two server farms. The server farms are under full administrative and physical control of NASS to maintain the agency's independence as a federal statistical agency. Each server farm hosts a cluster of virtualized servers. NASS also used virtualization software to virtualize the servers, thereby consolidating 94 servers into 44 virtualized servers in the two locations. Each of the two server farms was set up to service all NASS employees in case one server farm becomes unavailable due to an unexpected incident, such as a power failure caused by a storm.

The most critical component of the centralized environment is telecommunications. NASS implemented three tiers of telecommunications in each Field Office – primary, backup, and disaster recovery. The primary tier is two T1 lines trunked together. A T1 line is a leased fiber optic internet line that carries about 60 times more data than a normal residential modem. The backup tier is a Digital Subscriber Line (DSL). The disaster recovery tier is a wireless hot spot that can connect up to five desktops in a Field Office. If a Field Office loses its primary connection, the backup connection automatically comes online. NASS installed optimization equipment to reduce bandwidth consumption by compressing traffic and caching. NASS also purchased software to monitor the bandwidth utilization so that any stress on the network can be detected before becoming a serious problem.

Two storage area networks or SANs serve as dedicated networks for storage and access to data and applications, as shown in Figure 1. The SANs provide high-speed shared storage for the virtual servers and desktops. Each SAN has several terabytes of storage with additional shelves for expansion, which is sufficient disk space for the NASS environment.

NASS implemented redundancy in all critical technology components by providing redundant servers, storage, databases, switches, and telecommunications. In the event of failure of one server farm, employees can be connected to the other server farm. NASS is automating the disaster recovery and failover processes and will use load balancing to automatically shift employees to the other server farm in the event of a failure at a location.

The first two Field Offices were migrated to the centralized environment as a proof of concept in April 2010. NASS then conducted a pilot migration with two additional Field Offices in July 2010. An external, independent verification and validation of the centralized environment was performed by experts in the field of virtualization after the first four Field Offices were migrated. NASS then implemented the recommendations from this evaluation and began migrating additional Field Offices in October 2010. A NASS technical team with Headquarters and Field Office members went to each office to perform the transformation. The migrations of all 46 Field Offices and Headquarters were completed by September 2011. Different components of the virtualized network will be reviewed and upgraded each year to continually improve the network's performance. Hardware will be replenished on a four- to five-year cycle.

Since this transformational initiative has been completed, the cost savings can be measured. The cost efficiencies come from staff, hardware, and energy savings, but mostly from staff savings. Prior to this transformation, almost all Field Offices had a person serving part-time as the LAN administrator for the office. Through virtualization, NASS saved 23 full-time employee equivalents or \$1.9 million in annual salaries and benefits for Field Office LAN administrators. The annual savings in hardware is about \$350,000 or about five percent of the agency's information technology budget for hardware, software, and services. The software costs are approximately the same before and after virtualization. Prior to the virtualization initiative, NASS Field Office servers operated at about 25 percent utilization. Integrating the servers through virtualization resulted in energy cost savings since servers are now operating at 90 percent utilization centrally. The exact energy cost savings are not known, as the electrical costs were not available for all Field Offices.

In 2012, NASS will spend approximately $980,000 in salaries and benefits for information technology specialists to administer the virtualized network (desktop, server, storage, and telecommunications administration). The average annual cost for the network's hardware and software is about $250,000.

3.2. Standardize Metadata and Integrate Data

The focus of this transformational effort was to move to a centralized survey processing environment that uses consistent metadata from design through dissemination and enterprise databases instead of the decentralized architecture using proprietary data files, such as Blaise, SAS, and FoxPro (Nealon 2010). NASS was committed to this effort not only to minimize work inefficiencies caused by the decentralized data management process, but also to improve survey methods.

An example of minimizing work inefficiencies using a centralized database is eliminating the need to create, manipulate, and transfer thousands of Blaise, SAS, and FoxPro files each year. An example of improving survey methods using a centralized database follows. Many of NASS' surveys are repetitive so interviewers contact the same farm operator often within and across years. Unfortunately, a wealth of previous survey data has been historically underutilized in the survey process due to the decentralized data management environment. Therefore, integrating previous survey and census data in a centralized database that is readily accessible by NASS employees provides opportunities to improve survey methods by making better use of previous survey and census data from farm operators. Sampling and estimation procedures can be enhanced, survey management decisions can be better informed, edit and imputation processes improved, analytical capabilities broadened, and data products for data users expanded by making fuller use of previous survey and census data.

The standard descriptive survey metadata at NASS is provided in a single metadata database. There are three levels of standardized metadata in the new architecture: enterprise, survey-specific, and state-specific metadata. The enterprise metadata is metadata that must be consistent across all surveys, such as master variable names for survey questions. Some metadata may be survey specific, such as survey time (year, month, or week) and survey keying codes for data entry. Finally, some metadata can be unique to a state, such as conversion rates or rounding rules for commodities.

Fortunately, NASS had considerable expertise in database design and several successful enterprise database implementations prior to committing to transform all data processing to centralized databases. NASS has deployed two types of database designs depending upon the survey function (Kimball 1996): Online Transactional Processing (OLTP) and Online Analytical Processing (OLAP). OLTP databases are characterized by a large number of short online transactions (retrieve, insert, update, and delete). The main emphasis is on transaction speed and data consistency. To achieve this objective, the database design usually consists of a highly normalized data model, often with many tables, and focuses on optimizing single record processing time. The model is normalized so that any addition, deletion, or modification to a particular field is made in just one database table (Silberschatz et al. 2002). An example of a previously successful OLTP database at NASS is the Survey Questions database. NASS selected an OLTP database software for new database development that satisfied their transactional processing demands, was easy to administer, and was low cost. NASS has successfully been using this software for OLTP processing since the 2007 Census of Agriculture and has saved over one million dollar in software license costs.

On the other hand, OLAP databases are often characterized by retrieving a large number of records for aggregation, analysis, and developing reports based on the analysis. The main emphasis is on retrieval speed of multiple records. To achieve this objective, the database is typically designed with many fewer tables than an OLTP-designed database in order to optimize access speed to multiple records. NASS uses a dimensional database design referred to as a star schema for its OLAP databases (Kimball 1996). A dimensional database design or star schema consists of a central data table that contains all the data, which is joined to metadata or dimension tables that describe the data, such as metadata tables describing the *what* (survey variables), *when* (year and month), *where* (state and county), and *who* (information on each farm operator) for each data value in the data table. The NASS Survey Responses Data Warehouse, which contains over nine billion historical survey and census responses from farmers and ranchers in a seven-table database design, is an example of an OLAP database. Usage statistics for this Data Warehouse are impressive and show that over 95 percent of the queries from employees resolve in less than a few seconds. NASS has used a specialized database product for OLAP since 1998. This database product has been easy to administer and has surpassed NASS' expectations for retrieving data in a very timely manner. The dimensional design of the OLAP databases has also been easy for employees to understand and access.

Figure 2 provides a graphical depiction of the NASS enterprise databases. The enterprise databases are hosted on a UNIX/Linux platform, which has provided the performance, scalability, and security needed for earlier successful deployments at NASS. NASS also administers and controls the UNIX/Linux servers as an independent federal statistical agency. The centralized databases already in production have a gray background in Figure 2. The databases that have been developed and are in the pilot phase have the lined background. Finally, the two centralized databases with the white background in Figure 2 were being developed during 2012.

A brief description of the purpose of each centralized database follows.

3.2.1. OLTP Databases

The *Metadata* database is used to provide standard metadata descriptions, such as variable names and keying codes. The *Survey Questions* database contains all survey questions used in NASS surveys and is the source for survey questions for paper and pencil interviewing (PAPI), computer-assisted web interviewing (CAWI), computer-assisted personal interviewing (CAPI), and soon for Blaise computer-assisted telephone interviewing (CATI). When the Metadata database is updated, the relevant updates are automatically posted to the Survey Questions database. The *Sampling Frames* database contains individual farm, ranch, and agribusiness information used for sampling and

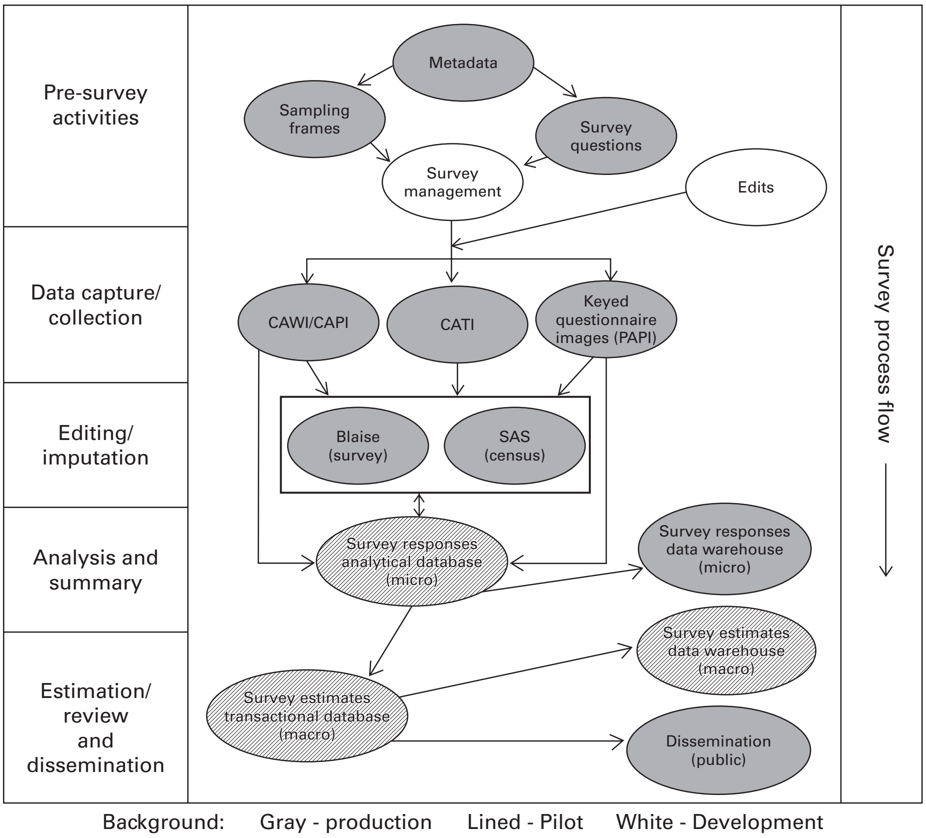

Fig. 2. NASS Enterprise Databases for Survey and Census Processing

survey purposes. The *Survey Management* database focuses on survey preparation and management activities. For example, information on each individual in the sample for a survey is provided, such as personal identifiers, sampling weights, and the mode of data collection. The Survey Management database and application will facilitate conducting multi-mode surveys efficiently. As an example, the system will allow switching data collection modes from mail only to CATI or CAPI late in the survey period to improve response rates. Furthermore, if a particular region, state, or county has a lower than acceptable response rate, while other areas have already surpassed their acceptable response rate, CATI interviewers can be readily redirected to contact the areas in need. The *Edits* database was in development during 2012 and will store edit logic and edit limits needed for CAWI, CATI, CAPI, and any edit or imputation application.

There are three *Data Capture* centralized databases – one for survey data captured from about 250 surveys annually over the web using CAWI or CAPI, one for data captured through the Blaise data collection software due to the special database requirements of Blaise Version 4, and one containing images of each page of a survey questionnaire for verifying data visually during the data edit and analysis phases. Data collected using the Blaise software was transformed from a distributed data environment using Blaise data sets for each survey in each Field Office to a centralized Data Capture database. When Statistics Netherlands releases Blaise Version 5, NASS will explore the possibility of integrating the CAWI/CAPI and CATI databases into a single database.

Data collected through CAWI, CATI, CAPI, and PAPI are then stored in a centralized *Editing/Imputation* database for the data editing and imputation process. A different editing/imputation database is currently used for the SAS editing/imputation application developed for the Census of Agriculture than for surveys using the Blaise edit software, because more database tables are needed for the Census of Agriculture. Finally, the *Survey Estimates* transactional database contains survey estimates, such as the corn for grain yield in Iowa as of September 1, 2012, that are calculated in the summary/tabulation application for survey variables. This database also contains available administrative data. The survey estimates and administrative data, if available, are reviewed and analyzed to derive published agricultural estimates.

3.2.2. OLAP Databases

The *Survey Response*s analytical database is a near real-time database containing multiple years of data in addition to the current survey's data. This database provides the current and historical data for data analysis and summary/tabulation processes, and also contains information needed to administer surveys (such as check-in information), track the disposition of collected data, and generate management information reports. The Survey Responses analytical database will also contain multiple years of information about farm operators on their response patterns, for example, what time of day they tend to respond and what data collection mode they tend to use. This information can be used to target farm operators more effectively when and how they prefer to be contacted, which could potentially improve response rates.

After a survey is completed and the agricultural statistics have been disseminated to the public, the micro-level data are loaded to the *Survey Responses* Data Warehouse. This database contains the deep history of farm, ranch, and agribusiness data and contains survey and census data from 1997 to the present. Similarly, there is a *Survey Estimates* Data Warehouse, which is the analytical database containing all historical survey estimates and administrative data for additional analysis and research purposes by NASS employees. Finally, the public-facing *Dissemination* database includes the U.S. published agricultural estimates or statistics and serves as the official data source for all public data products. This integrated database consists of only five tables and uses hierarchical metadata for easy browsing and data access, as discussed by Lehman and Nichols (2010). With hierarchical or nested metadata, a one to many relationship exists between metadata elements. The hierarchical metadata design is shown at http://quickstats.nass.usda.gov for the Dissemination database. The hierarchy allows, for example, a data user to select "vegetables" and then see all commodities associated with vegetables. In closing, a one table view of the Dissemination database was also designed for ease of use by NASS employees who prefer to navigate one rather than five tables when querying the database.

It is too early in the transformation process to quantify the overall staff resource savings from eliminating the work inefficiencies that existed in the decentralized data environment, since NASS is in various stages of the database deployments. NASS management plans to document the impact on staff resource utilization after a new centralized database has been in production for an entire annual cycle of surveys. However, four examples of cost savings already realized will now be discussed:

(1) ***Survey Questions Database:*** The implementation of the Survey Questions system (database and application) saved two of ten full-time positions in the Questionnaire Design Section at NASS while also simultaneously providing the ability to generate paper and web questionnaire instruments.
(2) ***CATI and Editing Databases:*** Staff savings of about six full time employees, mostly Field Office staff, were saved when CATI and editing in Blaise migrated to a centralized database environment, since the operational execution and support were simplified.
(3) ***Survey Responses Analytical Database:*** NASS documented that the seven mathematical statisticians responsible for determining the probability samples for NASS surveys spend approximately one-third of their time appending previous survey data to the many SAS data files for sample operators. This survey task will not be necessary when the Survey Responses analytical database is in production, which will save about two staff positions in this area.
(4) ***Dissemination Database:*** The new Dissemination database automated the loading process from the Survey Estimates Transactional database to the Dissemination database so the staff resources needed to load the published agricultural estimates from over 400 reports annually is less than .25 of a full time employee. The previous Dissemination data tables often required manual intervention during the loading process and required almost 1.5 full time employees.

3.3. Database Optimized, Generalized Survey Applications

The objective of this transformational initiative was to develop generalized application services that are thin-client and operate optimally with the enterprise databases. Examples of NASS applications that had to be re-engineered to operate optimally with the centralized databases are Blaise CATI and interactive SAS data analysis. The priorities for the application development were as follows. Highest priority was developing the database optimized, generalized application services, starting with the survey functions NASS planned to centralize or regionalize first, such as list sampling frame maintenance and Blaise CATI. The second priority was developing the generalized application services needed to standardize the 150 nonstandard surveys in the Field Offices. Finally, the third priority was retiring duplicative applications. As an example of retiring a duplicative application, NASS has used two CATI applications (Blaise and an in-house developed application) and one in-house developed CAWI/CAPI application. Due to cost considerations and the need to use in-house and contractor application development resources more effectively, NASS decided to commit to Blaise as its sole CATI application. In addition, NASS will explore the use of Blaise Version 5, when it is released by Statistics Netherlands, as a potential replacement for NASS' in-house CAWI/CAPI application.

In the past, NASS had a tendency to develop multifunctional applications that were complex, challenging to maintain, and not easily replaced. For this initiative, NASS developed generalized application services that were simple functional modules, such as an imputation service and an analysis service. By developing simple application services or modules, it is easier to add or enhance the services later. Also, a simple application service has an application programming interface (API) that allows other applications to use it. With this change in approach, NASS transitioned away from using the waterfall application development approach, which usually requires significant up-front planning, specifications, and design due to the complexity of the multi-functional application. NASS transitioned to the agile application development approach since a simple application service often requires significantly less up-front planning, specifications, and design than a multifunctional application (U.S. Government Accountability Office 2012). In addition, the agile approach delivers the application service in iterations and thus calls for continual communications between statisticians and developers, which has promoted more cohesive teamwork at NASS. The agile approach reduces the development time due to the reduced requirements for a simple application service, and it facilitates the re-use of existing application services by other applications.

NASS only had a few employees with advanced skills in both thin-client application development and databases. Therefore, NASS partnered with three application development contractor firms for most of the application development needed for this initiative. NASS also explored the use of off-the-shelf software for some application services.

The challenge with developing generalized applications was not with finding someone to develop the applications, since application development contractors are plentiful. The challenge was with finding contractors to develop applications that fully satisfied the business needs, were fully integrated with the agency's infrastructure, made optimal use of the databases, and satisfied the security standards of a federal statistical agency. McDonough (2010, p. 1) wrote: "Asking why major systems fail with regularity is like asking why there are no more .400 hitters in Major League Baseball. Success is simply too difficult." For example, many application development projects fail before a line of code is written, primarily due to improper business requirements gathering. NASS believed that it needed to implement an innovative governance process to maximize the likelihood of successful application development efforts. Therefore, in addition to applying standard project management principles, such as prioritizing application development efforts, managing scope creep, defining achievable deadlines, facilitating effective communications and collaboration, and monitoring costs, NASS implemented the following ten-step governance process:

(1) ***Senior Project Manager:*** All application development efforts were assigned to a senior executive, who served full time as senior project manager for the three transformational initiatives, was attached to the NASS Administrator, and reported weekly to the senior executive team to ensure any barriers to success could be promptly addressed. The senior executive set the vision, strategy, and priorities with input from employees who would play critical roles in advancing the development and implementation of the database optimized, generalized application services. The senior executive was the agency's contract officer for all application development contracts and had oversight for the initiation, planning, execution, and implementation of 21 application development projects. The senior executive formulated the budget for each project and could recommend the cancellation or

redirection of any project to the NASS Administrator. The NASS Administrator was responsible for allocating the required funds. As senior project manager, the senior executive was also responsible for managing the budget, ensuring deadlines were met, and collaborating with the senior executive team to select the project managers for each project.

(2) ***Project Managers:*** A statistician was assigned as project manager for each of the 21 application development projects to ensure the business needs were fully satisfied. The project manager was devoted full time or part time to the project, depending upon the scope of the development. The project manager also recommended team members with the appropriate skill sets to the senior executive team for approval. The project manager closely monitored the project for compliance in meeting the required specifications, and elevated any concerns to the senior project manager for resolution. Information technology specialists also partnered with the project managers throughout the project.

(3) ***Business Requirements Teams:*** A team of statisticians, led by the project manager, worked closely with the contractor and NASS information technology specialists to develop the business requirements for each application developed. NASS contracted a project management training company to train project managers on gathering and documenting user requirements.

(4) ***Enterprise Architecture Council:*** The NASS Enterprise Architecture Council (NEAC) provided documentation on the agency's technical standards for applications, databases, infrastructure, and security to the application developers. The NEAC reviewed the proposed technical solution from the developers to ensure it was aligned with the agency's technical environment. The project manager for each application being developed worked closely with the NEAC throughout the application development process in case any technical issues surfaced. Any compliance concerns from the NEAC were elevated to the senior project manager for resolution.

(5) ***Application Shadowing Teams:*** For each application development project, two members of NASS' Application Services Branch were assigned to shadow the contractor, review the code throughout the development, provide input, and familiarize themselves with the application so there could be a smooth transition from development to testing to production.

(6) ***Contractors:*** The contractor also assigned a project manager for each application development project, who documented biweekly any issues that might have surfaced so they could be resolved in a timely manner. The NASS senior project manager also met with a contractor's project manager biweekly to ensure the projects were progressing as expected and to resolve any issues.

(7) ***Security Reviews:*** NASS conducted security reviews and vulnerability scans of each application during development to ensure a secure application was being developed.

(8) ***Application Testing:*** The project manager led a team of employees from Headquarters and the Field Offices who performed thorough user acceptance testing of an application before giving approval to move the application into production. Payment for services was not made to a contractor until the project manager was satisfied that the application was ready for production.

(9) ***Change Control Board:*** A final technical review was performed by the NASS Change Control Board, in collaboration with the NEAC, before scheduling a generalized application service to be moved into production. This review ensured all technical components – security, databases, infrastructure, and applications – were ready for deployment.

(10) ***Support Services Contracts:*** NASS partnered with the contractors on support services contracts during the first few months of deployment to provide a smooth transition into production and maintenance.

Figure 3 provides a summary of the generalized application services used by NASS for the entire survey process. The generalized applications shown with the gray background in Figure 3 are already database optimized and in production. Generalized applications shown in the lined background are in production, but are not using the centralized databases optimally yet. Finally, the four generalized applications with the white background in Figure 3 were being developed in 2012. Since the generalized applications are being developed as simple services and not as complex, multi-functional applications, this will simplify improving or replacing various application services on an ongoing basis rather than needing to pursue major redesigns of complex multifunctional applications in the future.

As an example of a successful deployment of a generalized application service, go to http://quickstats.nass.usda.gov to use the Dissemination application and database (called Quick Stats) to access U.S. published agricultural statistics. Only two of the generalized application services have been challenging to advance into production. The Key From

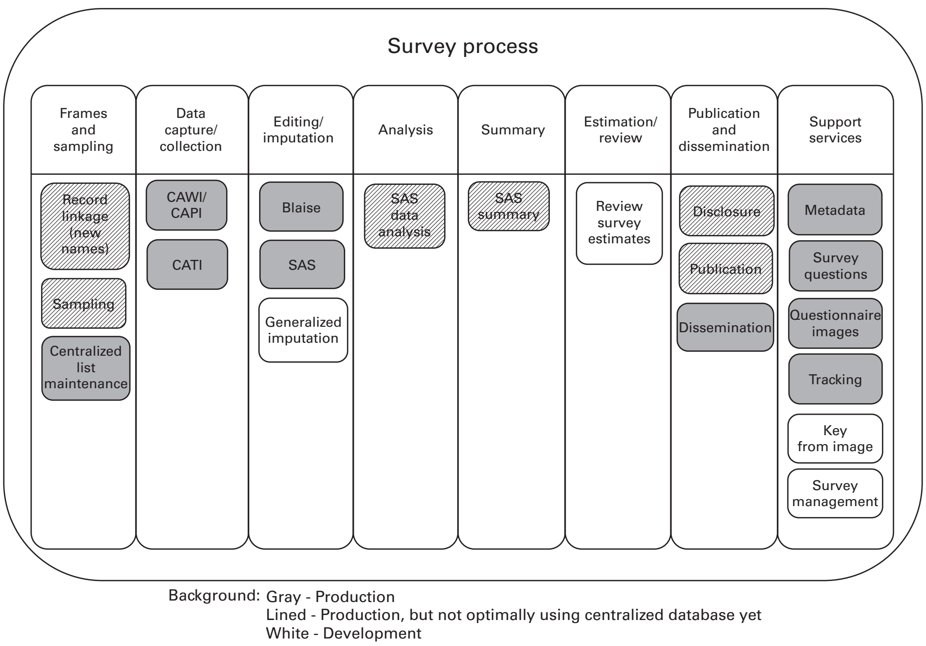

Fig. 3. NASS Generalized Application Services

Image application has not been deployed yet, mainly because of communication challenges between the NASS project team and the off-site contractor that resulted in a misunderstanding of some important business requirements. The Review Survey Estimates application is behind schedule primarily due to the testing and deployment of the new application competing for the same staff resources currently using and maintaining the legacy system.

It is too early in the application reengineering process to quantify the overall staff resource savings from developing the database optimized, generalized application services. However, staff savings are expected, for example, when the 150 nonstandard surveys are converted to the transformed architecture.

Metrics will be used to measure the effectiveness of the aforementioned application services. For example, deploying Blaise CATI in a centralized database will allow for measures, such as call attempts per hour and completed interviews per hour, to be readily available for all interviewers across calling sites. The Tracking application will allow for analysis of measures such as the count of questionnaires processed, overall and by work unit, and the time elapsed for various processing steps to be completed. Measures from the Key From Image application, including keying speed and number of errors, will be tracked. Measures from the Centralized List Maintenance application will provide the time for a request to be processed and the requests completed per hour by each list sampling frame maintenance employee.

With proper measures in place, the applications can be further enhanced to improve upon any work inefficiencies or error-prone steps. In addition, allocation of staff resources to survey activities can be improved based on the objective metrics. A NASS goal is that dissemination due dates of agricultural statistics will be more easily attained or shortened due to more efficient survey processing.

4. Lessons Learned

There were a number of key factors that contributed to the successful execution of the three transformational initiatives. The support from the NASS Administrator was unswerving and sufficient resources were allocated to the initiatives. A business case, vision, and implementation plan were documented and shared with all employees. Priorities were clearly articulated and followed. Experienced and competent project managers were selected to lead the various projects within the initiatives. Project teams consisted of skilled employees. Sound project management principles were followed. Consultants and contractors were utilized well to fill the voids in expertise at NASS. Efforts were made to keep all employees informed and prepare them for the technology changes. For example, monthly written progress reports on the transformational initiatives were shared with all employees. The factor that contributed the most to success was that the employees on the project teams worked well together, devoted many extra hours to the initiatives, and excelled in their responsibilities.

On the other hand, NASS feels the execution of the initiatives could have been improved in the following ten ways:

(1) ***Senior Executive Support:*** Progress was hampered at times when one or more of the ten NASS senior executives were not fully supportive of a project. Unity at the top

from the senior executives in an organization is critical to success, so try to get all senior executives on board throughout a transformation.

(2) ***Continuous Improvement:*** A transformed system, such as the centralized and consolidated network, should not be viewed as completed when initially deployed. An initial deployment should be viewed as the initial version of a system, and the agency should continue to improve the system over its life cycle. For example, NASS had more down time and slower connection time with the virtualized network than desired after the initial deployment, which could have been minimized if the enhancements requested at the initial deployment had been pursued in a timely manner. So the challenge is convincing the senior executive or executives with funding authority for the initiative that the transformation is a journey, not a destination, so that adequate resources are invested for enhancements throughout the life cycle.

(3) ***Employee Involvement:*** An effort was made to involve many employees from across the agency in the three transformational initiatives to obtain diverse perspectives and build support. As an example, LAN administrators from multiple Field Offices served on the team to consolidate the network. This helped Field Office employees adjust to the migration, since they were often more comfortable when colleagues from other Field Offices were part of the migration effort. In addition, considerable input was obtained from subject matter specialists in NASS when formulating the technical designs and implementation plans for the three initiatives, which resulted in their support from the beginning of the initiatives. However, there were still more opportunities for NASS to involve employees in the initiatives. Stewart (2012, p. 60) wrote: “When you choose to involve your people in decisions that affect them, they will feel valued because they know their opinions matter. They will also be far more inclined to support the final decision because they feel a sense of ownership.” Therefore, more involvement of employees in a transformation initiative can lead to more support from employees, which can result in a more successful project.

(4) ***Communications:*** The senior project manager should interact as much as possible with the project managers in a “face-to-face” manner to build trust, develop and strengthen relationships, and promote effective communications. These face-to-face communications will not only help the senior project manager better understand any issues that need to be resolved from the project teams, but will also provide opportunities to discuss the short-term goals and long-term direction for the initiatives with project managers to solidify a shared vision.

(5) ***On-Site Contractors:*** NASS was fortunate to have excellent contractor assistance for most of the projects. NASS did experience challenges with one contractor firm working off site, as the communications were not effective when trying to resolve problems in a timely manner. NASS recommends that the contractors work on site with the employees to foster excellent communications and collaboration and be able to resolve issues in a timely manner.

(6) ***Commitment to Solution:*** On a few occasions during the transformations, someone would suggest an alternative software or hardware solution be considered. It is critically important to embrace the proposed technical solutions, if they are working very effectively, until the initial implementation has been completed so that unnecessary delays are avoided.

(7) ***Design for High Availability:*** When the servers and data were distributed throughout the Field Offices, NASS had 46 points of failure. However, a failure, such as a server crashing, would only affect one Field Office. With the centralized design, the infrastructure and data are concentrated in two locations, so an outage is a high-impact event affecting about half the agency. Therefore, it is important to build in redundancy for all critical components to minimize the unavailability of the network.
(8) ***More Training:*** NASS wished it had allocated more resources for training employees. For example, more training would have been necessary for the network operations and maintenance staff on the virtualized network environment prior to full deployment so they could have supported the new environment more effectively. More training on using the centralized databases was needed for employees. With more training, NASS believes the transitions to production would have been smoother.
(9) ***More Shadowing:*** Even though a formal shadowing process was set up between the contractor and the NASS employees who would eventually be responsible for an application, the employees sometimes did not take full advantage of the shadowing opportunity so the support of the application was not as effective initially. NASS is more closely monitoring the shadowing process for the applications still under development to ensure more effective support is provided when deployed.
(10) ***Timelines:*** Project teams are often optimistic about how long it will take to complete a task, such as the time needed to gather business requirements or thoroughly test an application. It is suggested that a project team consider adding at least 25 percent to the proposed completion time estimate to arrive at a more realistic completion time.

5. Consolidation and Standardization of Survey Operations

NASS was finally in a position to explore the consolidation and standardization of survey operations to save staff resource expenses since the technical barriers were being removed by the three transformational initiatives. In addition, the hope is that the quality of survey operations will improve over time through consolidation and standardization. NASS is already pursuing two major consolidation and standardization efforts, which will now be discussed.

5.1. Centralization of Several Survey Functions

The first consolidation and standardization effort was to centralize several survey functions at a National Operations Center (NOC) opened in October 2011 by NASS in St. Louis, Missouri. This center would operate using the centralized network, enterprise databases, and generalized application services. The NOC is the culmination of the need for a more efficient centralized survey processing model that would reduce the redundancy of some survey functions across the 46 Field Offices and potentially improve data quality through the use of standardized procedures. As stated by the Secretary of Agriculture (Vilsack 2012): “With this centralized facility we are streamlining government operations and delivering results – at lower cost – for the American people, and in particular the people who provide and use our agricultural data products and services.”

The initial focus of consolidation and standardization at the NOC is on the following four survey functions that historically have been performed in most of the Field Offices: list sampling frame maintenance, paper questionnaire processing, telephone interviewing, and preparing training materials for NASDA interviewers. A brief summary of each effort follows. NASS cannot provide accurate measures yet of staff savings from these consolidation efforts, since the efforts have not been fully deployed.

5.1.1. List Sampling Frame Maintenance

Historically, employees in the 46 Field Offices maintained the list sampling frame of farmers and ranchers. The NOC has started to perform list sampling frame maintenance for all Field Offices, with employees at the NOC assigned to different geographic regions in the country to develop the necessary knowledge about the region to properly service the Field Offices in the region. Field Offices can submit proposed list frame actions to the NOC, such as name updates and additions, but the NOC performs the list frame maintenance using the centralized list frame maintenance application shown earlier in Figure 3. The consolidation and standardization of the list sampling frame activities should provide more consistent list sampling frame maintenance decisions than when list sampling frame maintenance decisions were made in each of the 46 Field Offices. The goal is that the centralized process will preserve or improve the coverage of the list sampling frames, but whether this will succeed is not known at this time since the centralized list sampling frame maintenance process has not been fully implemented yet. In the past, the Field Offices generally took great ownership of and pride in trying to improve the list sampling frame for their particular state, so a similar commitment will be needed by the NOC to preserve or improve the quality of the list sampling frame.

5.1.2. Paper Questionnaire Processing

Traditionally, the Field Offices received paper questionnaires from farm operators and NASDA field interviewers and then keyed the data. The NOC has started to receive, check in, create questionnaire images, key, and track all completed mail questionnaires from the Field Offices using generalized application services, such as the Tracking application, and the centralized databases. In addition, NASS is implementing CAPI across the country with NASDA field interviewers for face-to-face interviews and some remote CATI. NASS is leveraging its CAWI system, private cloud technology, broadband transmission, and the use of iPads for data entry and collection without storing any data on the CAPI iPad (Kleweno and Hird 2012). This provides an opportunity to reduce the cost of mailing questionnaires from field interviewers to the NOC, reduce the time lag to make the data collected available in the centralized databases, and reduce nonresponse rates since field interviewers can visit or call sampled farm operators until the last data collection day for a survey.

5.1.3. Telephone Interviewing

In 2009, NASS had six Data Calling Centers (DCCs), which each conducted the telephone interviews for two or three Field Offices and 29 separate Field Offices that did telephone interviewing for their own Field Office. The NOC is now serving as a Data Calling Center and has the capacity for 154 concurrent telephone interviewers. The NOC is also responsible for scheduling the work for the other six DCCs and maximizing the telephone data collection response rate for assigned surveys across the calling centers. Telephone calling from the 29 Field Offices was stopped by the end of February 2012 so that telephone calling would occur in seven rather than 35 locations. This will provide opportunities for more standardized interviewing procedures. It is not known yet if the benefits from more standardized interviewing procedures will eventually outweigh the benefits of interviewers from the 29 Field Offices, where telephone interviewers were often knowledgeable not only about the local agriculture, which can be useful when conducting interviews, but also local situations, such as weather and local agricultural events, which can affect a respondent's availability.

5.1.4. Preparing Training Materials for NASDA Interviewers

Previously, most training materials for NASDA telephone and field interviewers were developed and maintained in each Field Office. NASS plans to develop consistent training materials in the future at the NOC for use by all NASDA telephone and field interviewers to promote more standardized interviewing procedures in the DCCs and in the field. A Survey Statistician in a Field Office will continue to provide the training locally to the NASDA field interviewers using the standard training materials, complemented with information relevant to local agricultural situations.

The NOC opened in October 2011 with a very small staff and has been adding employees during 2012. Meaningful measures to assess the impact of these four consolidation and standardization efforts, such as the impact on the list sampling frame coverage or telephone response rates or survey costs, will not be available until the NOC is in full production mode. During the transition phase in 2012, NASS has experienced a decline in response rates for national surveys, such as the March 2012 Crops/Stocks Survey, the August 2012 Agricultural Yield Survey, and the September 2012 Crops/Stocks Survey. The hope is that this trend will be reversed when the hiring and retention of telephone interviewers stabilizes in the seven DCCs and when CAPI for field interviewing and remote CATI is fully implemented in the Field Offices to complement the efforts at the DCCs.

A quality assurance program is being developed at the NOC to provide quality metrics in a dashboard for various survey activities. The first implemented component of the program is a Telephone Interviewer Evaluation Form where information gathered from monitoring telephone interviewers is presented to supervisors immediately after completing a monitoring session so that any corrective action can be taken. The quality assurance program being developed at the NOC will eventually be implemented also in the other six DCCs.

NASS is in the initial stage of testing Computer Audio Recorded Interviewing (CARI) at the NOC. CARI records the telephone interviews, which will provide insights into respondent and interviewer interactions. NASS will be able to review and rate the recordings to use as a coaching tool to improve the performance of interviewers. CARI will also provide NASS with the ability to pretest questionnaires.

5.2. Consolidation of Field Offices

The second area of consolidation and standardization is NASS' proposed realignment of its 46 Field Offices into Regional Field Offices and State Field Offices. A Regional Field

Office is proposed to service multiple Field Offices and focus on the data review process (editing and analysis) and the coordination of reimbursable surveys for the states assigned to the region. Field Offices not serving as a Regional Field Office will remain viable as a State Field Office and are proposed to be staffed with a State Statistician and a Survey Statistician. The State Statistician will primarily manage cooperative agreements with local agricultural organizations, the State Department of Agriculture, and land-grant universities. In addition, they will maintain relationships with the local agricultural industry and interact with farmers to promote survey cooperation. As mentioned previously, the Survey Statistician will serve as the Agency's point of contact with the NASDA field interviewers in the state, and will work with employees from the NOC to provide consistent survey training to the NASDA field interviewers in the state.

The consolidation of Field Offices will help NASS manage its survey program with fewer employees. After the consolidation, statistical measures will be needed to determine if the cost savings have been accompanied by an increase or decrease in the quality of the survey program, for example, the refusal rate and the inaccessible rate increased or decreased for surveys.

Finally, NASS realized that the success of consolidation efforts depends partially on effective communications among the NOC, DCCs, Regional Field Offices, State Field Offices, and Headquarters. Therefore, NASS implemented Video Teleconferencing (VTC) throughout the agency (Ballou 2009). VTC combines audio and video to provide a means for communication, collaboration, and decision-making in real time among multiple sites. VTC enables every participant to see each other's facial expressions and body language, and allows people to share files and data, in order to present information, review documents, and make decisions. VTC is being used to hold meetings and conduct training. The initial implementation costs for hardware, software, staff time, bandwidth fees, and contracting services have been offset by travel savings for employees. In a time of tighter budgets, travel costs have been saved by using VTC instead of transporting employees physically to various meeting or training sites.

In closing, the consolidation and standardization efforts being pursued will save staff resource costs to position NASS to continue to provide valuable data products and services in service to U.S. agriculture in an era of tighter budgets. The impact of consolidation and standardization initiatives on the quality of the NASS survey operations is not known at this time and needs to be measured, understood, and reacted to so that the quality of the survey program is hopefully preserved or improved in the future. Also, NASS should closely monitor the impact of the consolidation initiatives on the morale of employees, who are the agency's best asset and are often coping with the changes the consolidation efforts bring, such as the need to relocate to a Regional Field Office.

6. Summary and Discussion

NASS has made excellent progress advancing the three architectural transformations that were prerequisites for consolidating and standardizing survey operations. The centralized and virtualized network has replaced the decentralized network from 48 locations. Standardized metadata and centralized databases were put into production for many survey functions in December 2012. The development of database optimized, generalized application services has progressed with several of the application services already in production and the remaining planned for production during 2013.

The most critical keys to success for the three transformational initiatives were the NASS Administrator's support, the provision of sufficient resources, and the effectiveness of the project teams. The key to longer-term success for the transformed architecture is the NASS senior executive team committing the necessary resources annually to continually improve the network, metadata and databases, and generalized application services so they properly serve the evolving needs of employees in the future. Collins (2001, p. 162) said: "How a company reacts to technological change is a good indicator of its inner drive for greatness versus mediocrity. Great companies respond with thoughtfulness and creativity, driven by a compulsion to turn unrealized potential into results." The unrealized potential of the transformational initiatives is to improve the quality of the survey operations and therefore the quality of data products NASS provides to the public.

In the past, NASS has leveraged the 46 Field Offices and local NASDA interviewers to perform critical survey functions, such as building relations with local data providers to encourage survey cooperation and identifying farm operators not on the list sampling frame. NASS has started to centralize and regionalize survey functions in response to tighter federal budgets. A future mandate for NASS is to measure objectively the impact of the consolidation and standardization efforts on the quality of the survey operations, such as measuring the increase or decrease in list sampling frame coverage and survey response rates due to consolidation and standardization. Only then will NASS know if the consolidation and standardization efforts not only saved staff resource costs, but also preserved or positively impacted upon the quality of the survey operations.

7. References

Ballou, H. (2009). Project Implementation Plan – Video Conferencing. National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C. Available at: http://www.nass.usda.gov/Research_and_Science/Technology/ Video%20Conferencing%20Plan.pdf (accessed March 12, 2012).

Collins, J. (2001). Good to Great: Why Some Companies Make the Leap . . . and Others Don't. New York: HarperCollins Publishers.

Gleaton, E. (2011). Transformational Initiative #1: Centralizing LAN Services – Vision and Implementation Plan. Washington, D.C.: National Agricultural Statistics Service, U.S. Department of Agriculture. Available at: http://www.nass.usda.gov/Research_and_ Science/Technology/Centralizing%20LAN%20Vision%20and%20Implementation%20Plan.pdf (accessed February 20, 2012).

Kimball, R. (1996). The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses (First edition). Hoboken, NJ: John Wiley & Sons, Inc.

Kleweno, D. and Hird, P. (2012). New Solutions, Challenges, and Opportunities: CAPI the NASS Way. Conference of European Statisticians, October 31–November 2, 2012, Geneva, Switzerland. Available at: http://www.nass.usda.gov/Research_and_Science/Technology/CAPI%20the%20NASS%20Way_New%20Solutions%20Challenges%20and%20Opportunities.pdf (accessed September 26, 2012).

Lehman, J. and Nichols, M. (2010). Data Warehousing and Web 2.0 Data Dissemination, Fifth International Conference on Agricultural Statistics, Uganda. Available at: http://typo3.fao.org/fileadmin/templates/ess/documents/meetings_and_workshops/ICAS5/PDF/ICASV_2.3_049_Paper_Lehman.pdf (accessed April 17, 2012).

McDonough, F.A. (2010). 6 Keys to Avoiding Another ‘Grand Design’ Failure. Federal Computer Week. Available at: http://fcw.com/articles/2010/10/25/comment-frank-mcdonough-grand-design-program-management.aspx (accessed May 7, 2011).

Nealon, J. (2010). Transformational Initiative #2: Database-Optimized, Generalized, and Modular Applications (DOGMA) – Vision and Implementation Plan. National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C. Available at: http://www.nass.usda.gov/Research_and_Science/Technology/Database%20Optimized%20Generalized%20Applications.pdf (accessed April 17, 2012).

Silberschatz, A., Korth, H., and Sudarshan, S. (2002). Database System Concepts, (Fourth edition). New York: McGraw-Hill.

Stewart, J.P. (2012). 52 Leadership Gems: Practical and Quick Insights For Leading Others. Lake Oswego, OR: J.P. Stewart Systems, Inc. Published by Leadership Excellence at www.LeaderExcel.com.

U.S. Government Accountability Office (2012). Software Development: Effective Practices and Federal Challenges in Applying Agile Methods. Report GAO-12–681, July. Available at: http://www.gao.gov/assets/600/593091.pdf (accessed August 15, 2012).

Vilsack, T. (2012). National Association of State Departments of Agriculture Winter Policy Conference. Available at: http://mfile3.akamai.com/23747/wmv/ocbmtcmedia.download.akamai.com /23747/nass/020712.asx (accessed July 14, 2012).

VMWARE (2006). On-demand Webcast: What is Virtualization Technology and Why You Should Care. Available at: http://www.techrepublic.com/webcasts/on-demand-webcast-what-is-virtualization-technology-and-why-you-should-care/4410685 (accessed May 7, 2008).

Received June 2012
Revised September 2012
Accepted October 2012

Journal of Official Statistics, Vol. 29, No. 1, 2013, pp. 29–48

The General Survey System Initiative at RTI International: An Integrated System for the Collection and Management of Survey Data

Lisa Thalji[1], *Craig A. Hill*[2], *Susan Mitchell*[2], *R. Suresh*[2], *Howard Speizer*[2], *and Daniel Pratt*[2]

There is a risk of introducing survey error at every stage of any study involving a survey: design, data collection, processing, and analysis. Effectively managing the survey sample from the time of sample selection through the survey lifecycle is essential to producing high-quality data on schedule and within budget. Managing the survey lifecycle using software systems that are not fully integrated can result in error and cost inefficiencies. The development of an integrated data collection and management system that supports monitoring of survey error has the potential to reduce errors and improve operational efficiencies. This system, referred to as Nirvana, uses a standardized database, protocol, and terminology. It integrates case status and history information across modes of data collection and tracing as well as sample and contact information. Nirvana also centralizes questionnaire development, quality monitoring, and support for real-time survey management decisions.

Key words: Total survey error; data collection systems; Adaptive Total Design; paradata; survey costs.

1. Introduction

Organizations that collect, process, and analyze survey data must design and implement systems and protocols that guard against the intrusion of survey error to produce data and statistics that are of sufficient quality. However, quality is a complex construct. By some definitions, quality data must be accurate and meet the user's needs or fitness for use (Juran and Gryna 1980). Biemer and Lyberg (2003) notes that when there are a variety of uses for survey data, fitness for use becomes multifaceted. He stipulates that survey data must be accurate, credible, comparable, useable, relevant, accessible, timely, complete, and coherent (Biemer 2010). In other work, this Total Survey Quality framework has been described as having three dimensions: accuracy, timeliness, and accessibility (Biemer and Lyberg 2003). In this article, we describe RTI International's (RTI) initiative to build an integrated survey management system designed to improve the accuracy, timeliness, and accessibility of the data we produce.

RTI is a nonprofit research organization that primarily conducts research for the U.S. federal government. As a research contractor, RTI responds to requests for proposals from

[1] RTI International 3040 East Cornwallis Road, Post Office Box 12194, Research Triangle Park, NC 27709-2194, U.S.A. Email: thalji@rti.org
[2] RTI International, 3040 East Cornwallis Road, Post Office Box 12194, Research Triangle Park, NC 27709-2194, U.S.A.

U.S. statistical agencies as well as other nonstatistical agencies. Research proposals are requested in response to policy or program questions (e.g., understanding incidence and prevalence rates for drug use and abuse) and the scope of research services requested varies widely. Some agencies request study designs and plans for data collection, data processing, and statistical analyses; others require only data collection and data processing proposals. RTI conducts studies using all types of survey modes, subject populations, and establishments. In all proposals, RTI describes how it will collect the relevant data and, most importantly, how it plans to ensure the quality of those data given schedule and budget constraints.

There is potential for introducing survey error at each stage of a survey – from sampling frame creation to data dissemination (United Nations 1999). Sampling error arises from observing a sample rather than the whole population, and further errors can be introduced through deficiencies in the sample frame or sample design. Nonsampling error can be introduced by questionnaire authors, interviewers, and/or respondents (measurement error); errors result from bias in the collected data (nonresponse error); and errors can occur in the coding, editing, and management of collected survey data (processing error). The accumulation of error across all survey stages is referred to as *total survey error* (Biemer and Lyberg 2003). Any error at any stage diminishes the accuracy and precision of survey estimates.

One common approach to minimizing survey error is to use, for each stage of a survey, a software-based solution designed to double-check, prevent, or alert survey managers to the potential for error. In the data collection process, for example, survey software can automatically disallow entry of out-of-range values, check for (and disallow) entries that are inconsistent with prior answers, and can keep interviewers from asking questions that are inappropriate for certain subsets of respondents (such as asking male respondents if they are pregnant). Another common approach is to use software to collect and analyze paradata (i.e., data about the data) during the data collection process that allows survey managers to adjust resources and approaches in near-real time to gain improvements in survey quality metrics, such as response rates, while balancing time and cost constraints. Paradata can range from simple summary measures such as response rates to something as specific as the elapsed time between call attempts.

At RTI, we call the process of using the paradata to make management choices that improve the quality of the survey data Adaptive Total Design (ATD). Evolving from the concept of responsive design (Groves and Heeringa 2006), ATD is a framework for monitoring key statistics that help survey managers decide how best, in real time, to modify protocols to minimize total survey error and maximize cost efficiency (Hunter et al. 2012). RTI's ATD system requires access to all of the data collected during survey administration and complex models for measuring the impact of these protocol changes on survey estimates.

Before this initiative, RTI built project-specific and independent software, systems, and protocols designed to prevent and minimize error. This process resulted in separate software-based systems for managing sampling, data collection, coding, editing, and data analysis processes. The information contained in these systems was not integrated (i.e., the systems did not talk to one another) within a given survey or across survey modes.

This lack of integration created the potential for undetected measurement and data processing error.

In addition, as is the case at many large survey research organizations, many of the survey management systems were developed for particular projects, which meant that RTI often had multiple software systems to support the data collection stages of the survey process. Since these systems were built independently of one another, standardization was inherently lacking, which also could create the potential for specification error. Finally, because many of these systems were built some time ago, it was proving increasingly difficult to generate the kind of near-real time paradata needed to fully use our ATD system.

In the data collection and data processing stage especially, we saw a need to create an integrated and standard system that ties together all stages of data collection and data processing by employing a common database structure, protocols and terminology across studies. Consequently, we developed an RTI-funded initiative, the General Survey System Initiative (GSSI). The GSSI's main goals are to improve the quality of the data we produce through means not achieved heretofore with our custom approach: (1) integration, (2) standardization, and (3) accessibility to necessary paradata.

The GSSI is governed by a group of senior managers in RTI's survey and programming groups. The senior managers are responsible for carrying out various aspects of the initiative, including technical design, budget and schedule.

2. Process for a New Approach

Our focus on customization had introduced several operational challenges. Although study staff were able to develop workarounds, concern grew that the lack of standardization and integration and the misallocation of resources could be causing unintended and undetected survey error. For example, if the tracing system and computer-assisted telephone interviewing (CATI) system are not integrated, a separate process is required to share information about the sample member between the two systems. The time delay to associate the sample member's information from two different systems impacts the survey manager's abilities to take immediate action and array resources appropriately. It also may result in some duplicative systems and/or manual operations in an attempt to improve the timeliness of information and decision making, thereby affecting overall operational efficiencies and costs. These types of challenges and inefficiencies made it increasingly difficult to maximize data quality within reasonable costs (i.e., within the budget provided by the government) as our survey efforts grew more and more complex.

To understand the challenges of working with RTI's systems, the GSSI steering committee decided to gather information by surveying RTI project leaders, survey managers, and programmers. A web survey was developed to (1) inventory all of the software and related tools used to support surveys; (2) evaluate the efficiency of deploying these tools; and (3) evaluate the overall usability of software and related tools. These areas typically contribute 20–30% of the total survey costs. Over the span of several months, the data were mined for common themes and challenges. The committee methodically reviewed and discussed these data to ensure a shared understanding of each problem before proposing solutions. What emerged through these discussions was a clearer understanding that the root cause of our problems could be attributed to the lack of standardization, integration of survey systems, and access to survey data and paradata (Couper 1998).

Another consideration that factored into our discussions was the need to better position ourselves to address the changes and challenges in the field of survey research. Response rates have been declining (see, e.g., Curtin et al. 2005; de Heer 1999; Steeh et al. 2001; Tortora 2004) across all modes of data collection: in-person, telephone, and mail surveys. To counter this trend, the conduct of multimode surveys has increased as a means of obtaining higher response rates than those using one mode alone. Often, telephone surveys are combined with an in-person follow-up. In other instances, in-person surveys are combined with shorter telephone surveys and/or mail surveys to bolster the response rates. To continue to provide timely quality data in a fast-changing survey environment, we needed robust systems that could operate under various conditions and provide the type of paradata required to efficiently manage survey operations.

Table 1 summarizes our root cause analysis and serves as the basis for developing a new approach to address these objectives. These objectives are guided by our total quality management framework and desire to improve the accuracy of the data we collect by reducing and controlling nonresponse error, measurement error, and data processing error.

The next step in our process was to convene a multidisciplinary task force to explore the feasibility of building or buying a single, unified data collection system to manage surveys across data collection modes and activities. The task force developed a list of general and function-specific requirements to guide the evaluation. General requirements included the following:

- Integrate case management systems across all modes of data collection and tracing.
- Integrate software systems for questionnaire development.
- Integrate data quality metrics.
- Store survey and paradata for all data collection modes and tracing in a single database.
- Standardize event and status codes to conform to survey industry practices for reporting and tracking survey progress.
- Allow for easy configuration to support a wide range of studies.
- Provide for easy-to-use tools to access paradata for survey management.

The task force evaluated several commercial off-the-shelf (COTS) data collection software systems. Where possible, questionnaires were programmed and case management systems were tested in the COTS software. The questionnaire programming systems met our requirements, but the case management systems did not. Given the level of complexity and details required to effectively manage the types of large-scale surveys RTI conducts for the U.S. federal government, a robust and refinable case management system was deemed the most important criterion. Thus, after careful consideration of these COTS data collection software systems, RTI decided to invest overhead funding to develop its own integrated data collection system to maximize design control over the integration of the case management systems. In making this decision, we knew that a commitment to such an endeavor was a long-term investment that would involve staff participation across disciplines for a two- to three-year period, followed by ongoing enhancements.

Table 1. Objectives for new approach

Objectives	Challenges	Desired outcomes
Standardization	Custom approach causes inefficiencies Maintenance of multiple systems raises costs Diversity of systems requires additional training and knowledge transfer	Eliminate duplicative systems Improve operational efficiencies Decrease learning curve for users and programmers Reduce level-of-effort for system specification, setup, and support Provide more flexible staffing options Leverage reusable components of the system Increase consistency of implementation
Integration	Difficulties in identifying next actions for particular cases result in duplicative work Delays in sharing information across systems reduce efficiency and increase cost	Reduce elapsed case processing time Eliminate case processing steps Improve sample management information Eliminate redundant data Reduce reliance on programmer intervention to move information between systems
Access to data (survey and paradata)	Lack of real-time access to data increases difficulty and cost of survey management Changing priorities quickly during data collection is difficult without access to flexible, dynamic data reporting tools	Increase availability of data to survey managers Improve quality of decisions Improve timeliness of decisions Develop better and more consistent data access and reporting tools

Once this decision was made, the GSSI steering committee convened another task force whose mission was to develop the ideal structure, flow, and functionality of a unified data collection system that would meet our objectives of standardization, integration of survey systems, and access to survey data and paradata. Their initial focus was limited to the data collection process, thereby addressing the most immediate costs and quality concerns and challenges of operating in our current environment. On large-scale surveys, data collection costs can be as high as 50–60% of the total costs of the survey, and can contribute to errors in the accuracy of the data collected. The design for the unified data collection system resulted in the development of the Nirvana system.

3. Design and Architecture of Nirvana

The integrated data collection system – Nirvana – is designed around a central database and repository. Data are stored in the standardized database in a table structure and are accessible by all of the component systems that support case flow as the sample moves through different stages of activities. Each project using Nirvana has its own database that holds all data, including:

- sample data, which includes strata used for sample selection and other characteristics of the population of interest.
- subject contact data, which includes contact information as well as past activities specific to each sample member.
- metadata that specifies details about the questionnaire instrument, such as question text, response options, and other attributes of the questions, as well as the configuration settings for the stages in the data collection process, and is comparable to the structural metadata used in Statistical Data and Metadata Exchange (SDMX).
- paradata which includes a wide array of operational and observational data about data collection, such as cost and production data, record of calls data, information about the sample frame and sample, data quality measures, interviewer characteristics, and interviewer observations, and
- survey data regardless of data collection mode.

The Nirvana database has a core SQL schema that holds all of the sample and subject contact data. All other modules tap these centralized schema tables to provide the necessary data for contacting sample members by telephone, by mail, or in person. When new contact information becomes available, these same tables are updated, regardless of the source of the information; thus, all of the modules receive the updated information at the same time.

Another key feature of the Nirvana database is that it maintains a single database table that holds the paradata, such as the history of all of the efforts expended in contacting the sample member. Thus, whether an invitation is sent via e-mail, telephone, or in-person visit to the sample member's home, these activities are recorded in the same table, and that information is available to all users. This feature ensures that the interviewer, supervisor, and survey manager are fully aware of all of the efforts made to reach the sample member.

Nirvana uses an administrative database that maintains linkage information in an encrypted form for all of the project-specific Nirvana databases. The administrative database also maintains the role and access control for each user for each project. When a Nirvana application or website is invoked, it queries the administrative database to ensure that the current user has access to the project in question and automatically connects to the desired project. This technique makes it easy for survey managers to access their projects and provides an extra level of data security for the projects. Nirvana design took into consideration resource allocation, especially developer resources, with the hope of reducing costs. For example, even though the initial setup of the database might require an experienced SQL programmer, their involvement will be minimal after the initial database setup phase. The configuration of the project and other setup work can be done by a programmer with less experience and hence at lower cost.

The database is set up in such a way that each survey process is defined as a *stage*; for example, a lead letter mailing; a telephone, field, or web-based interview; or the tracing or locating of sample members are each defined as a stage. The current state of each stage is reflected by the current *status* code for that stage. For any given case, a specific status code initializes the stage. The user posts an *event* to indicate that a specific activity had occurred, and the system automatically derives the new status based on the posted event, the prior status, and other indicators. Note that it is possible for a status code in one stage to trigger the initialization status code in another stage, which is a simple but powerful technique that is employed to seamlessly link different modes of data collection. The Nirvana events, status, and stages essentially codify the steps and process in the workflow of a project and to some extent automatically document the same.

For example, for an in-person study that requires confirming telephone and address information in an automated fashion as its first step, the survey sample is loaded into Nirvana tracing stage number 400 with a status code of 1001 to indicate that the tracing stage has been initialized. When sample cases are sent to tracing (an activity), the status code within that stage changes to 1101 to indicate that sample cases have been sent to tracing. The results from tracing are reflected in the status code of that stage. When all pre-field tracing activities are completed, the sample case moves to the next stage in the survey process: the lead letter mailing stage. The completion status of pre-field activities triggers the initialization of the lead letter mailing stage. As letters are mailed on a flow basis, the status code changes to lead letter mailed. The lead letter mailed status automatically triggers the initialization of the next stage, field interview, and so on throughout the data collection process. An advantage of this scheme is that at any time the user can easily review the history of a sample case and know which activities have been conducted to date. In addition, the user knows the total number of sample cases ever initiated at each stage. The above examples illustrate the adaptability of the Nirvana system to changes in workflow processes required by any survey project. A data collection or monitoring process can be described as a combination of actions (Nirvana events) and resulting steps (Nirvana status); this implies that the process can be defined as a Nirvana stage. In addition, if a sequence of tasks can be automated, a Nirvana module can be developed to standardize the implementation of those tasks for that project and then the module becomes re-usable for future projects.

Figure 1 depicts the data flow among the major components of Nirvana.

Case Management Systems. As an overarching system, Nirvana incorporates the previously separate case management systems for CATI, computer-assisted personal

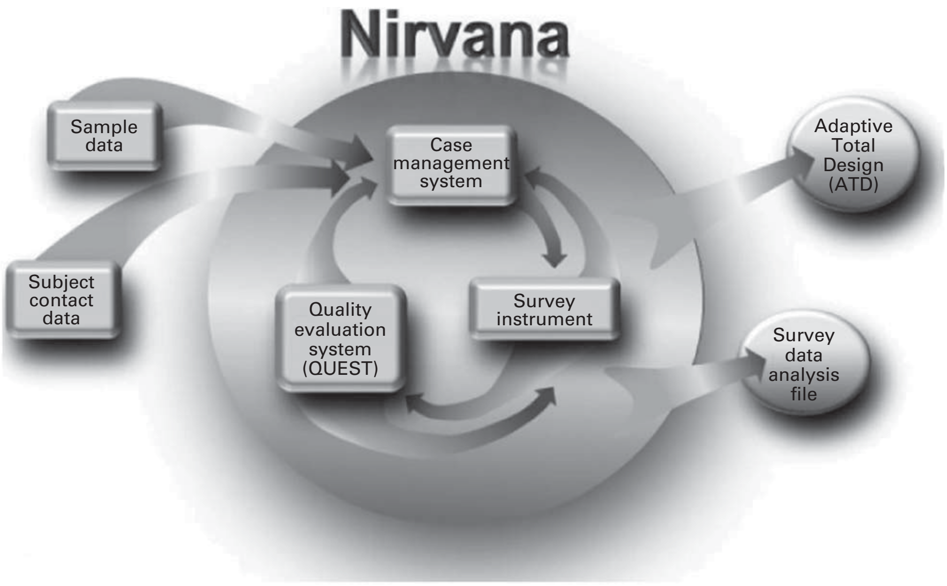

Fig. 1. Nirvana data flow

interviewing (CAPI), and tracing as well as modules that facilitate document mailing and receipt, e-mail invitations for web-based self-interviewing, biospecimen tracking, and so on. CATI and CAPI are for managing telephone and field data collection respectively; tracing is for identifying the best phone number or address for a sample member in a list survey or for panel maintenance. These case management systems and other modules use the same underlying tables to access the sample data and to record the paradata on the status of data collection. In some organizations, these systems use separate databases and processes that provide status information to a central database or reporting structure, but with Nirvana, the same database tracks the progress of data collection regardless of the mode, facilitating a seamless change in data collection mode. The integration of case management systems reduces programmer labor and minimizes the potential for processing errors. For example, if a sample member in a mixed-mode survey chooses to complete the case via the web rather than by CATI, the system will immediately detect that the survey was completed via the web so that no additional effort is expended in CATI for that sample member. Programmer involvement is not required to move completed surveys from one database to another, thereby decreasing the risk of processing errors.

Survey Instruments. The questionnaire development system is used to develop survey instruments for telephone, field, web, and hard copy paper data collection modes. This system allows for real time edit and consistency checks. Using the same system to author the instruments reduces training costs for the specification writers and programmers. This system provides special interfaces for translation specialists to develop instruments in other languages, including Spanish and Chinese. The system also facilitates recording of the questions for audio-computer-assisted self-interviewing instruments as well as mode-specific variations in question text and/or response options. The metadata for the survey instruments and survey data are stored in the Nirvana database. This facilitates the generation of codebooks to accompany analysis datasets as well as the generation of reports that can be stratified using key items in the questionnaire data.

Quality Evaluation System (QUEST). This component supports monitoring of interview quality for telephone or field data collection. QUEST addresses measurement error through its ability to listen to recorded excerpts of the questionnaire administration and provide immediate feedback to interviewers. With Nirvana, the QUEST tables are in the project database, and the results of the review can be easily combined with other information such as interviewer demographics, their caseload, and item nonresponse to provide a richer quality control report.

4. GSSI and ATD

Perhaps the most notable outgrowth of the GSSI is the enablement of ATD because this ties the initiative to the Total Quality Survey Framework that underpins RTI's approach to surveys. As described earlier, ATD is a survey management tool that monitors cost, productivity, and data quality on a real-time basis during data collection. It differs from electronic reports and quality checks because visual displays (i.e., charts, graphs, and dashboards) are used to exhibit quantitative information in multiple dimensions, including over time. ATD provides a snapshot of the progress and quality of a data collection that reveals deviations from expectations, prior observations, and protocols in a convenient and accessible way. It is designed to show and compare relationships and changes across data elements. While there are costs associated with its development and implementation, these costs can be offset by the reduced risk to data quality, and the ability to ward off problems that may be expensive to fix (for example, retrieving missing data or replacing interviews that were fabricated).

ATD is integral to the Total Quality Survey Framework because it exposes survey errors that are common across studies, as well as specific to a single study design. When survey errors are identified, survey managers intervene to reallocate resources or address problems before they become entrenched. For example, survey managers may shift interviewing resources from one sampling area to another to level response rates, retrain interviewers to remedy errors in interview administration, introduce changes to the data collection protocol (such as an alternate mode of data collection or an incentive) to increase response, or alter CATI programs to correct or prevent measurement or interviewer errors.

ATD development depended on an integrated system of data collection and management provided by the GSSI. First, paradata needed to be collected systematically and stored in a common format to enable adoption across a wide range of surveys; it also had to be linked to survey data. Data standardization and linkage were essential first steps. The ATD system was developed from the GSSI to be:

- the monitoring tool for total survey quality.
- the support for data-driven decision making.
- accessible by a wide range of project staff and clients, each with different informational needs.
- graphical in display but allowing ready access to underlying data and numeric reports.

- supportive of cross-survey methods research so that metadata could be created and analyzed to examine patterns across surveys.
- user-friendly and intuitive, giving end users the ability to create custom graphs and dashboards with minimal programming knowledge.
- accessible on a variety of platforms, including laptops, smartphones, and tablet computers, and
- affordable.

The ATD system arrays dashboards and charts under five topical headings: (1) production and cost data, (2) sample status, (3) record of calls data, (4) data quality, and (5) timing. Information about the interviewers – identification numbers, gender, age, years of experience, and education – is used to supplement and subset data in these categories, comparing, for example, the productivity of female interviewers with male interviewers. Survey managers have the ability to identify and investigate potential problems by examining related graphs and drilling down to the underlying data itself. Thus it is possible to view survey status and changes in both visual and numeric form. For example, Figure 2 shows the overall mean survey length by day and the interviewers who deviated by more than 1 standard deviation from the aggregate mean. The chart shows a fairly consistent length of about 60 minutes with convergence around the mean as the data collection period lengthens and interviewers gain experience. However, interviewer "Bill" exhibits a pattern of longer interviews, and interviewer "Cindy" exhibits a pattern of shorter interviews. These findings merit further investigation. In surveys with sensitive questions, for example, a longer administration time may mean that the interviewer is dwelling on questions that make the sample member uncomfortable, potentially leading to an interview break-off or item refusal. Survey managers have the ability to confirm findings by examining graphs showing the section timings of sensitive question groupings, as well as the underlying data showing the number of interviews Bill completed and his tenure as an interviewer. If the number of completed interviews is small (for example, one per shift) or Bill is inexperienced, the observed pattern may be anomalous and expected to change over time.

To illustrate other uses, charts can be displayed by sample geography so that resources can be redirected to areas where production is lagging. For example, Figure 3 displays

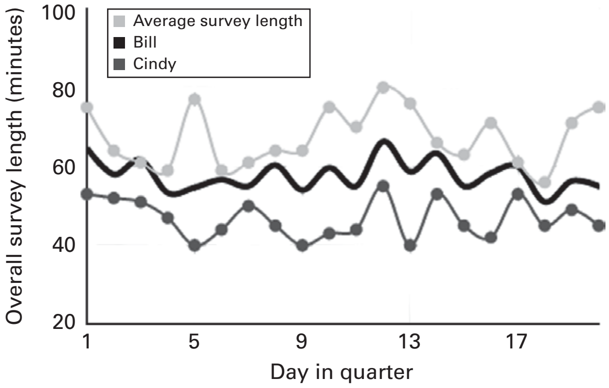

Fig. 2. Overall survey length and outlier interviewers

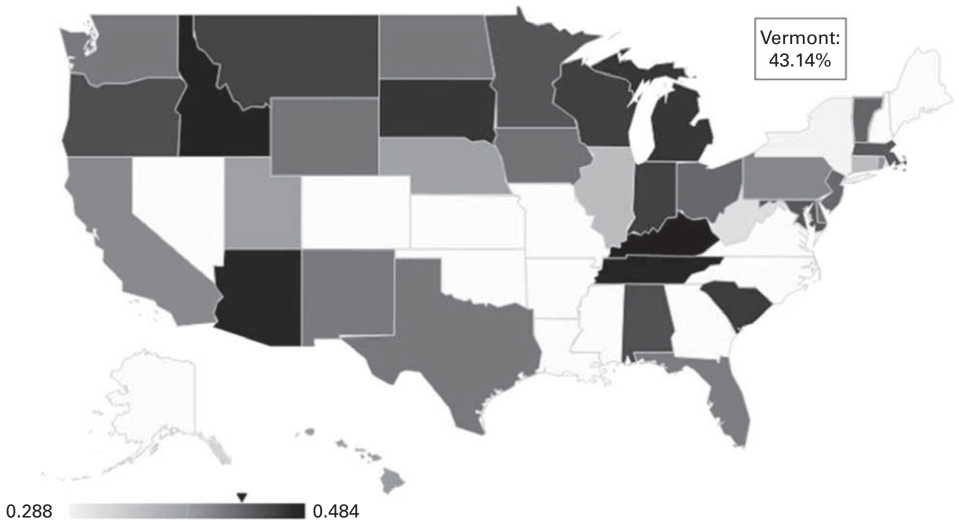

Fig. 3. Response rates by state

a map showing response rates by state for a national field study. Low-responding states can quickly be identified and investigated. Clusters of low-responding states can also be observed, indicating potential problems within regions that may be related to staffing or resources. By moving the cursor over the state, the actual response rate is displayed. As an enhancement, the percentage point improvement in response from the prior week or month can also be programmed to display.

Figure 4 is a chart showing the number of interviews completed by day compared with the 7-day moving average. The graph shows how production declines over the course of data collection, and how daily results differ from the moving average, indicating a more or

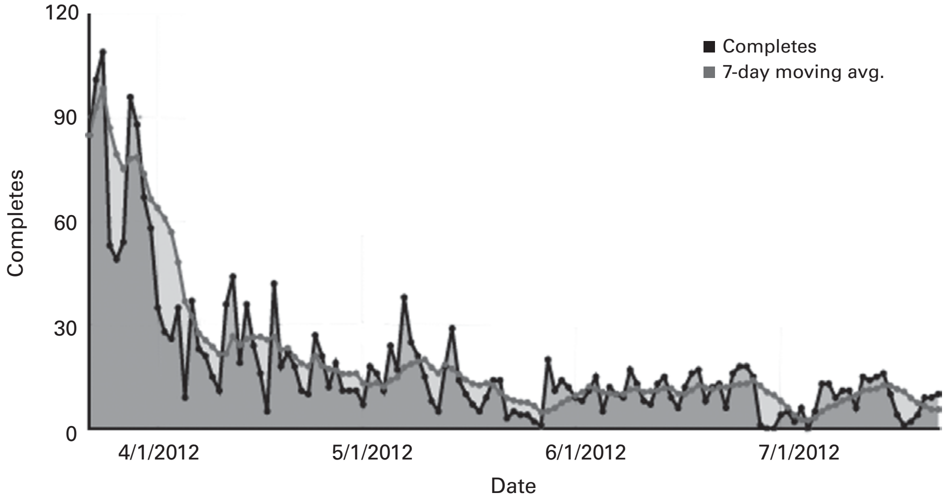

Fig. 4. Completed interviews by day compared with 7-day moving average

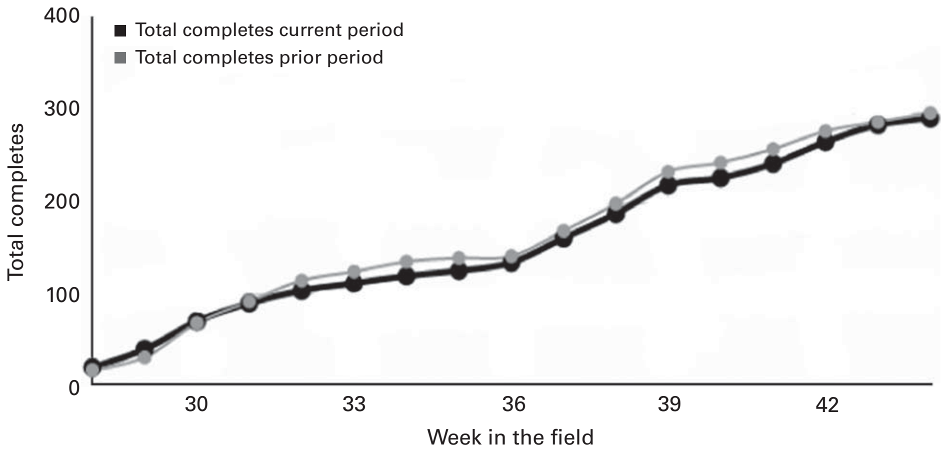

Fig. 5. Completed interviews by week compared with prior period

less productive day. Information in this chart can be used to model and predict final production, and to efficiently plan and staff future rounds of data collection.

In another view, the chart in Figure 5 shows how interview production in a current survey round compares to interview production in a previous round. In general, the chart shows that production increased at the same pace and rate in each round, which is evidence that field conditions remained stable and there were no external "shocks" to the environment that might impact data comparability. Expected production by week can be overlaid on the chart to compare planned and actual progress.

Because the charts incorporate the element of time, survey managers can immediately gauge the effect of changes to the data collection protocol or unanticipated field conditions. The chart in Figure 6 shows how the number of CATI interviews completed by

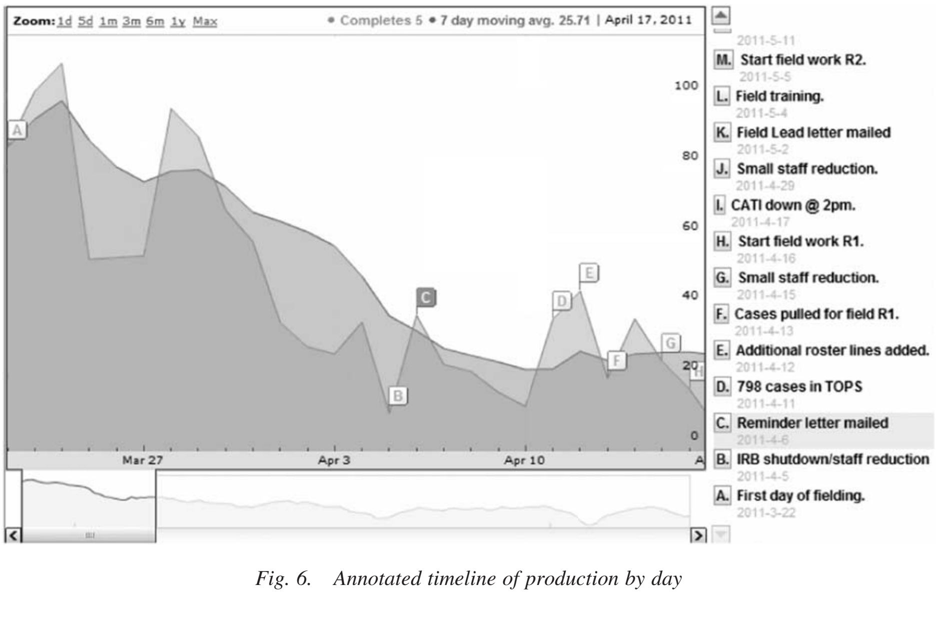

Fig. 6. Annotated timeline of production by day

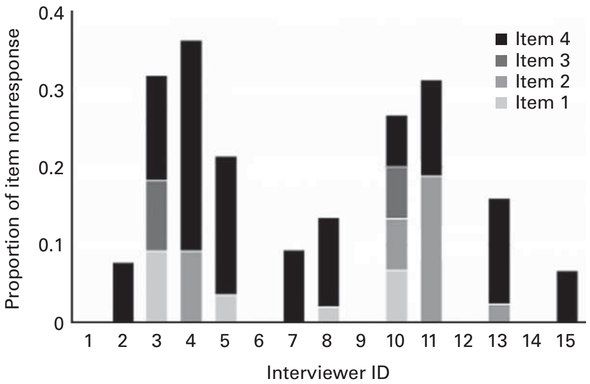

Fig. 7. Item nonresponse by interviewer

month tracked with interventions (reminder letters, incentives, sample releases) and external conditions (weather, early closings, system downtime). This chart provides a record of events that helps to explain the variation in production over time.

As an indicator of data quality, Figure 7 shows item nonresponse rates by interviewer for key items. It is possible to quickly discern which items are higher in nonresponse (e.g., Item 4) and which interviewers are experiencing item nonresponse at higher levels and across multiple items (e.g., Interviewers 3 and 10). This information can be used to coach interviewers to improve their performance. It can also be used to alert survey designers and analysts to items that may need imputation or re-evaluation for future use.

While cost and productivity measures are relatively limited and straightforward to chart and monitor, data quality measures can span the range of survey errors, including measurement errors, nonresponse errors, sampling errors, and processing errors. ATD has the potential to influence the occurrence and magnitude of all these error types, but is perhaps most practical for measurement error because it can be tied to paradata emanating from survey systems for reinterviewing, reviewing recorded interviews, and matching with administrative records – approaches commonly used for estimating measurement error. ATD is also well suited to identifying potential nonresponse error because it is capable of monitoring not only item nonresponse rates, but also survey nonresponse (weighted and unweighted) both overall and by key subgroups. ATD is best suited to applications where the underlying paradata exists in a standardized way in other survey systems, and when complex, study-specific programming (for example, the calculation of variances estimates) is not required.

To date, RTI has integrated ATD into a number of large-scale in-person and telephone surveys. In implementing ATD, the largest cost investment was in standardizing the data elements, and combining survey data and paradata in the same database, as described in Section 3. There were also costs associated with educating end users about ATD – the selection of relevant charts, a protocol for monitoring outcomes, and appropriate interventions to take when problems are discovered. For this reason, we found it cost-efficient to develop charts and graphs that could be used repeatedly over time or across multiple surveys. The creation of the graphs was straightforward, using commercially available charting software that did not require specialized training.

5. Implementation and Evaluation of Nirvana

Nirvana has been implemented with a phased approach. A "beta" or pilot version of Nirvana was implemented in the spring/summer of 2011 in the field test phase of a repeating cross-sectional education study: the field test of the National Postsecondary Student Aid Study of 2012 (NPSAS:12) conducted for the National Center for Education Statistics (NCES). The system was refined during this operational field test and rolled out on a second project in the summer/fall of 2011: the field test of the Baccalaureate and Beyond: 2008/12 Longitudinal Study, also conducted for the NCES. The ATD component was not available for this early evaluation period. Nirvana has been implemented in three full-scale studies in the first few months of 2012 with two more planned for later in 2012.

NPSAS:12, as a representative example of the projects handled by Nirvana, has multiple modes of data collection (web, telephone, in-person, paper) with many hierarchically linked survey participants (postsecondary institutions, students, and parents). The sample members are selected from school enrollment data and added to the survey on a flow basis. Survey data are collected over a nine-month period from over 100,000 students and their families and more than 1,500 postsecondary institutions. Data collection is monitored across all modes and by multiple subgroups. NPSAS:12 is used as the base year for a longitudinal study that follows first-time beginning students over time as they continue their education and work experiences.

The Nirvana system architecture, development environment, and software tools were finalized in the early stage of the NPSAS:12 field test implementation period. The critical component of this architecture was the database schema, including the metadata for all of the database variables. After finalizing the design and the database structure, the team produced a design document and scheduled a series of discussions with the system development team to form a common understanding of the technical design. The project manager established system development and change management protocols and assigned system modules for development to team members. The GSSI steering committee approved an implementation schedule with concrete resource requirements.

Although the system design decisions were made early in the implementation process, the data flow specifications were more difficult to finalize. As described earlier, the Nirvana system uses a three-tier case status hierarchy (i.e., stage, event, and status) to define case flow through all survey steps. Developing, standardizing, and mapping the three-tiered code structure to the American Association for Public Opinion Research codes was time intensive, but will yield benefits as all projects will report on the same standardized status codes.

The pilot projects generated a set of user-driven recommendations for enhancements to Nirvana and its associated data collection processes. The primary system implementation challenge, as demonstrated in the pilot studies and in later projects, has been to effectively automate the case flow protocol. Although Nirvana establishes a common case flow protocol, survey specifications for a particular project invariably require augmentations to the standard that need to be programmed and tested. As the system is enhanced to support multiple protocols, less augmentation and adaption is required, and this challenge is mitigated.

Conducting an evaluation of Nirvana is an integral part of our continued development of the system. An evaluation of the Nirvana system will assess the extent to which the objectives and desired outcomes have been reached with its implementation. The first 18 months of Nirvana roll-out entail a gradual ramp-up to ensure that various operational components are well-tested before a more widespread implementation. Our evaluation plan focuses on efforts during and after this roll-out period and consists of evaluating elements of each of the key objectives for Nirvana: standardization, integration of survey systems, and access to survey and paradata. The primary approach for the evaluation will be qualitative, though we will conduct descriptive analysis using some quantitative data available. For each of the key objectives, Table 2 specifies certain desired outcomes and the associated metrics included in the evaluation plan to assess the level to which those outcomes have been achieved.

To evaluate the standardization objective, RTI will track the number and types of projects and number of personnel that adopt and implement Nirvana. Although RTI senior management has the expectation that each new survey project will adopt and implement Nirvana, the extent to which this happens – and reasons for doing so or not doing so – will be monitored. For each implementation, we will track the set-up costs, including the time and effort to develop system specifications and to program and test all system modules. These costs will be compared against similar costs for prior projects that did not use Nirvana. Set-up costs will also be compared over time among projects implementing Nirvana. Also measured will be the extent to which one-time customization of processes is required for projects using Nirvana, with the goal that one-time customization is less necessary over time. Such customization will be measured within projects as well as across projects over time. Note that certain customizations will not be one-time-only efforts; they will comprise desired expansions of modules that will be made available for future projects – that is, incorporated into Nirvana's expanded capabilities. The evaluation necessarily will include qualitative as well as quantitative elements, given that no two projects are the same and there will not be directly comparable projects. In as much as quantitative data are available for comparisons and analysis, qualitative data obtained through focus groups will provide valuable context on several of the proposed outcomes. The perceptions of the end users regarding several of the proposed outcomes contribute to the determination of whether goals were achieved in implementing the Nirvana system. Focus groups will be used to engage in discussion and elicit feedback from various end users with varying roles across the projects.

Improving quality is a multidimensional goal that is part of each of the three objectives for building Nirvana. As such, measuring improvement in quality is an important component of our evaluation plan. The standardization and system integration objectives are designed to provide a common, easier-to-use framework for managing and tracking survey activities. Providing richer information to survey managers theoretically allows them to avoid unforeseen problems and proactively develop effective survey strategies. By achieving these objectives, we hope to significantly reduce the number of opportunities for mistakes (i.e., nonadherence to particular specified steps in the process) and consequently the number of errors committed by the survey personnel. Therefore, our plan is to measure quality improvement by tracking mistakes and, in particular, the amount of rework required by these mistakes. We plan to count mistakes that require repeating processes or correcting steps that have already been implemented and track the amount of labor that is invested in this rework. The amount of labor required to correct mistakes is a proxy

Table 2. Metrics for evaluating Nirvana

Objectives	Desired outcomes	Metrics
Standardization	Eliminate duplicative systems	Inventory systems deployed across projects
	Improve operational efficiencies	Count data repositories, files, and programs Count process steps
	Decrease learning curve for users and programmers	Assess extent of customisation required Evaluate usability
	Reduce level-of-effort for system setup, support, and maintenance	Assess cost and effort to specify, set up, and support the system
	Provide more flexible staffing options	Count the number of staff trained to set up Nirvana Count the number of staff that have used Nirvana to manage a survey
	Leverage reusable components of the system	Compare modules deployed across projects
	Increase consistency of implementation	Count the number of projects deploying Nirvana
Integration	Reduce elapsed case processing time	Assess case processing times for critical functions
	Eliminate case processing steps	Count processing steps
	Improve sample management information	Ask survey managers
	Eliminate redundant data	Count repositories, files, redundancies
	Reduce reliance on programmer intervention to move information between systems	Measure programmer time for system support
Access to data	Increase availability of data to survey managers	Ask survey managers
	Improve quality of decisions	Ask survey managers Ask clients
	Improve timeliness of decisions	Ask survey managers Ask clients
	Develop better and more consistent data access and reporting tools	Inventory reporting tools

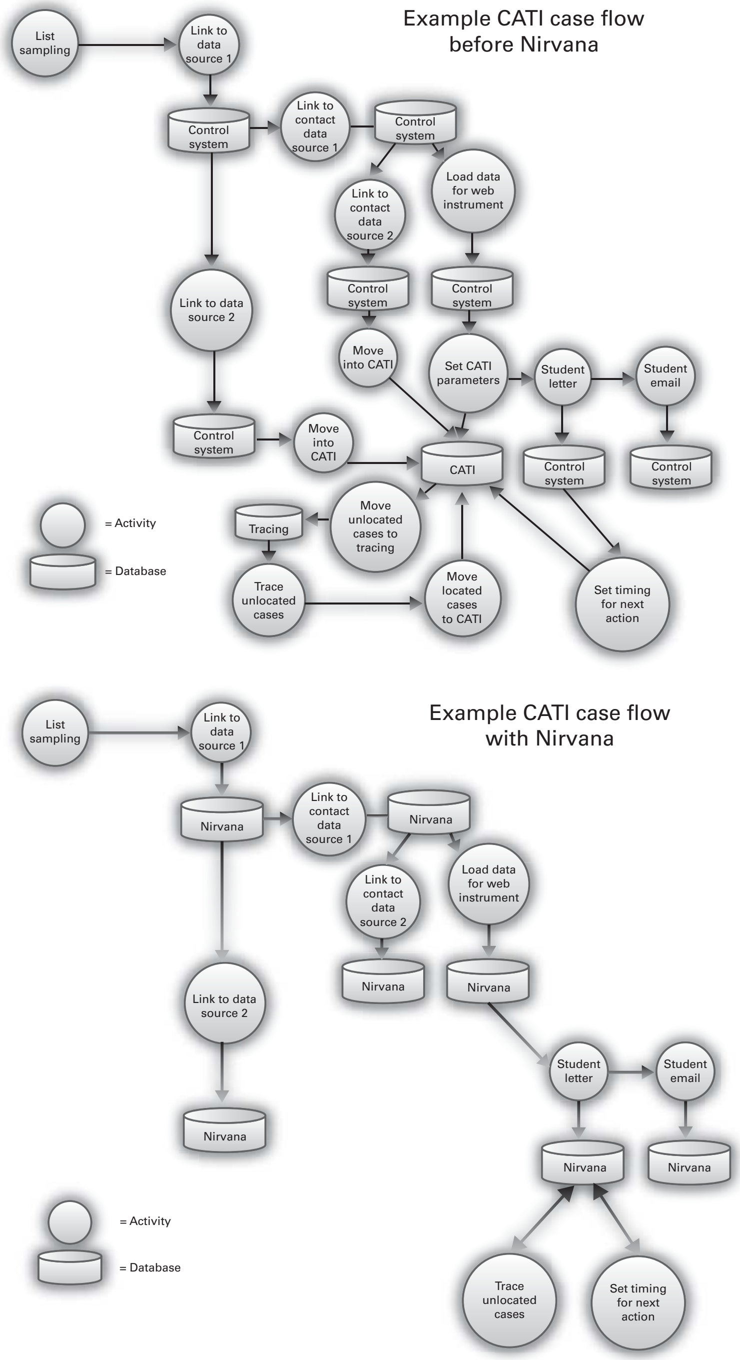

Fig. 8. Example process-flow diagrams

measure for the size of the mistake; therefore, with further use of the system we expect to see a lower mean mistake count and less effort invested to correct mistakes over time.

The system integration achieved through Nirvana addresses a number of important issues: reduction of steps in the process, reduction of the number of opportunities to make errors, increased efficiency, elimination of redundant data and workaround steps associated with multiple systems, and reduction of programmer intervention. Figure 8 depicts a pair of process-flow diagrams associated with the initiation of CATI data collection to illustrate this integration.

The first diagram shows the process flow for a portion of the survey process with pre-Nirvana systems, and the second diagram illustrates the flow for the same activities with Nirvana. A comparison of the diagrams shows a reduction in the number of activities with Nirvana, from 15 to ten. The number of databases has been reduced from three different databases (control system, CATI, and tracing) to a single database (Nirvana). Duplicative information has been eliminated, and information is shared across processes rather than requiring data transfers. The fewer steps and reduction to a single database in turn improve efficiency and reduce the risk of error associated with data inconsistency and redundancy.

A fundamental objective of Nirvana has been providing access to survey data and paradata that is easy, direct, and comprehensive for all who need it (e.g., project leaders, survey managers, data collectors, and programmers). A crucial step in meeting this objective is the establishment of a shared database across all system components and processes. Previously, there were multiple databases that had existed with RTI's data collection. Having all information across processes (e.g., sampling, mailing, tracing, self-administered data collection, CATI, and CAPI) in one location enables survey managers to make decisions in an informed and timely manner due to the availability of trustworthy, real-time status details. Making such data easy for survey managers to use is a work in progress with Nirvana. RTI is in the process of adding query and reporting tools to its ATD framework for Nirvana. To evaluate the success of these tools and access to data for informed decision making, RTI will include discussion of these topics in the focus groups of survey managers, programmers, and project leaders for each of the projects implementing Nirvana during the 18-month initial phase of implementation. Additionally, a user survey similar to the system inventory survey will be administered at the end of the same 18-month phase. An important aspect of this survey will be the extent to which survey managers had easy, timely access to key information to inform the appropriate next steps and to ensure the success of their studies.

6. Conclusions

The overarching goal of the GSSI is to improve survey quality by reducing errors and assisting in the production of relevant, timely, accessible, and cost-effective data. A root cause analysis was undertaken to better understand the challenges associated with a custom systems design and development approach to collect and produce accurate study estimates. This analysis uncovered that the lack of standardization and integration of data collection systems could result in unintended and undetected survey error. The analysis further revealed that the tools necessary to access survey and paradata easily and in real time to manage survey operations were not in place.

The three objectives – standardization, integration of survey systems, and access to survey data and paradata – were among the key design requirements for developing Nirvana, a unified data collection and management system. The standardization and system integration objectives are designed to provide a common easier-to-use framework for managing survey activities. The access to survey data and paradata objective is designed to provide direct and uncomplicated access to this information. Together, these objectives will provide better, timelier, and easier access to data that will enable survey managers to make more informed decisions regarding survey strategies and avoid more unforeseen problems. By achieving these objectives, our goal is to significantly reduce the number of opportunities for error and consequently the number of errors committed.

The Nirvana evaluation plan is designed to measure the extent to which each objective has been met and survey quality has been improved. Our plan details the outcomes expected and the process by which we will assess and collect further information. The plan includes the use of focus groups, surveys, quality improvement metrics, and cost estimates where feasible. Further enhancements will be made based on the results of our evaluation plan.

ATD is a noteworthy improvement to the survey process that was made possible by the integrated system provided by the GSSI. ATD maximizes survey quality by allocating budget and personnel resources to survey activities depending on priorities and the potential to reduce total survey error. It has been used to identify interviewers who deviate from an expected norm, triggering the need for further investigation or retraining. It has been used to monitor the status of the sample to determine the timing and size of sample releases. It has been used to monitor item nonresponse rates and response distributions to key survey variables. ATD has also been used to compare data collection costs with budgeted costs to identify discrepancies and forecast final costs. For example, in a recent longitudinal, in-person survey on child health and well-being, ATD charts showed that labor hours per completed interview were less than budgeted and expected to increase only slightly through the remaining data collection period. The anticipated cost savings were used to plan and fund a study on nonresponse bias in the survey estimates. The potential uses of ATD are plentiful: comparing weighted and unweighted response rates by sample geography over time, modeling final response rates, tracking indicators of bias, comparing characteristics of respondents with the sampling frame, identifying interviewer falsification, and observing the real-time effects of protocol changes such as contacts by different modes or the offer of incentives.

Together, GSSI and ATD have improved our ability to take advantage of previously untapped data to improve the quality of survey estimates. While time and resource intensive to establish, our investment is expected to pay off in terms of reduced costs and errors over multiple projects, greater operational efficiencies, and enriched information for decision making.

7. References

Biemer, P. (2010). Total Survey Error: Design, Implementation, and Evaluation. Public Opinion Quarterly, 74, 1–35.

Biemer, P.P. and Lyberg, L. (2003). Introduction to Survey Quality. New York: John Wiley & Sons.

Couper, M.P. (1998), Measuring Survey Quality in a CASIC Environment. Invited paper presented at the Joint Statistical Meetings of the American Statistical Association, Dallas, TX. Available at: http://www.amstat.org/sections/srms/Proceedings/ (accessed June 20, 2012)

Curtin, R., Presser, S., and Singer, E. (2005). Changes in Telephone Survey Nonresponse Over the Past Quarter Century. Public Opinion Quarterly, 69, 87–98.

de Heer, W. (1999). International Response Trends: Results of an International Survey. Journal of Official Statistics, 15, 129–142.

Groves, R. and Heeringa, S. (2006). Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs. Journal of the Royal Statistical Society Series A: Statistics in Society, 169, 439–457.

Hunter, D.C, Mitchell, S., Carley-Baxter, L., and Keating, M. (2012). Using Paradata to Monitor Survey Production, Cost, and Quality within an Adaptive Total Design Framework. Paper presented at the Federal Committee on Statistical Methodology Research Conference, Washington, DC. Available at: http://www.fcsm.gov/12papers/Hunter_2012FCSM_V-C.pdf (accessed June 20, 2012)

Juran, J.M. and Gryna, F.M. (1980). Quality Planning and Analysis: From Product Development Through Use. New York: McGraw-Hill.

Steeh, C., Kirgis, N., Cannon, B., and DeWitt, J. (2001). Are They Really as Bad as They Seem? Nonresponse Rates at the End of the Twentieth Century. Journal of Official Statistics, 17, 227–247.

Tortora, R.D. (2004). Response Trends in a National Random Digit Dial Survey. Metodolski zvezki, 1, 21–32.

United Nations (1999). Information Systems Architecture for National and International Statistical Offices – Guidelines and Recommendations. Available at: http://www.unece.org/stats/documents/information_systems_architecture/1.e.pdf (accessed June 20, 2012)

Received November 2011
Revised June 2012
Accepted October 2012

Journal of Official Statistics, Vol. 29, No. 1, 2013, pp. 49–71

Redesign of Statistics Production within an Architectural Framework: The Dutch Experience

Peter Struijs[1], *Astrea Camstra*[1], *Robbert Renssen*[1], *and Barteld Braaksma*[1]

In response to a changing environment, Statistics Netherlands has embarked on a large-scale redesign of the way statistics are produced. The aim is to increase the capability to respond to changing information demand, to lower the response burden for surveys, especially for businesses, and to improve efficiency, while preserving the overall quality level. The redesign is carried out within the framework of a so-called enterprise architecture, which gives overall guidance when structuring the processes of the organisation, including statistical methods and IT tools used. The article describes the redesign approach and explains the main features of the architecture. The emphasis is on experiences that may be relevant to other national statistical institutes operating in a similar environment.

Key words: Process design; enterprise architecture; IT expenditure; standardisation; GSBPM.

1. Introduction

The context in which official statistics are produced by national statistical institutes (NSIs) has profoundly changed in the last decades (Van der Veen 2007). Many NSIs are experiencing budgetary tightening and political pressure to reduce the administrative burden of survey participation, especially in the business sector (Ypma and Zeelenberg 2007). At the same time, NSIs are facing an ever-growing demand for statistical information. This relates to new and changing user needs for relevant statistics, for example about the financial downturn or immigration, as well as to the timeliness and amount of detail with which statistics are to be released. In the European Union (EU), the number of regulations requiring EU member states to produce statistics has multiplied in the last two decades, for instance in the area of short-term statistics and annual business statistics.

In addition, technological advances have resulted in more possibilities for survey and automated data collection, and administrative registers have increasingly become available for statistical purposes. The amount of data available for use as input for statistics has grown significantly and requires integrating different data sources in an intelligent way.

[1] Statistics Netherlands, Postbus 24500, 2490 HA, Den Haag, The Netherlands. Website www.cbs.nl, for all authors. Emails: psjs@cbs.nl, acma@cbs.nl, rrnn@cbs.nl, and bbka@cbs.nl, respectively.

Acknowledgments: The authors thank Geert Bruinooge, Dirkjan Beukenhorst, Frank Hofman and Kees Zeelenberg, Statistics Netherlands, for their contribution to the discussion and their comments on drafts of the article. Comments by the guest editors of JOS were especially valuable. The views expressed in this article are those of the authors and do not necessarily reflect the position of Statistics Netherlands.

To meet the challenges of a transforming environment, NSIs will have to adapt the way in which they produce statistics. This is widely acknowledged (Bae 2010; Braaksma et al. 2013; Doherty 2010; Falorsi et al. 2013; Galvin and Rhodes 2010; Lorenc et al. 2011; Studman 2010; Swiss Federal Statistical Office 2007). At Statistics Netherlands (SN), the following key strategic goals have been formulated to guide this transformation:

1. It must be possible to quickly and flexibly respond to changing information demand.
2. The existing high quality standards must be maintained and coherence improved.
3. The reporting burden should be minimal, especially for businesses.
4. The statistical processes must be cost efficient.

Although other NSIs face more or less the same challenges, they may choose to concentrate on somewhat different strategic goals. The strategic aims themselves are not the focus of this article, but rather the way in which they can be achieved. Translating the strategy into the necessary organisational changes is where an enterprise architecture can be of value.

Briefly, an enterprise architecture can be seen as a general blueprint of how to create an organisation's processes and IT infrastructure (Lankhorst 2009). It can serve as a guideline for the redesign projects that are carried out to bring about the changes as well as for the management of change. The Generic Statistical Business Process Model (GSBPM) is an international reference for describing statistical production processes used by NSIs at a high level (Vale 2009). However, this model only covers part of an enterprise architecture as it mainly focuses on statistical processes.

This article illustrates how the enterprise architecture developed and used by SN contributes to realising its strategic goals. SN began developing the architecture in 2004. The purpose of the article is not to give a historic account, but to identify factors, pitfalls and practices that may be relevant to other NSIs operating in a similar environment and facing similar challenges. Obviously, the enterprise architecture has been developed given a specific context. SN is a centrally organised institute that is responsible for providing 95% of official statistics in the Netherlands. This situation is quite different from that in other countries where there are several producers of official statistics like ministries and regional agencies. Institutional independence is guaranteed by law. European statistical requirements have a serious effect on the work programme of SN: more or less 70% is determined by European regulations, which are legally binding for the EU member states. Nevertheless, most key elements of the enterprise architecture – and the architectural approach itself – are independent of the local context.

We start in Section 2 with arguing for an integral approach to producing statistics, moving away from the traditional approach of independent, parallel production lines (sometimes called 'stovepipes' or 'silos'). The integral approach is incorporated in the enterprise architecture of SN, which is described in Section 3, together with an example of its application in economic statistics. The section illustrates how the architecture makes the strategic goals operational. In particular, it shows how the architecture can help to increase the responsiveness to new information needs and reduce the response burden. Having an enterprise architecture alone is not sufficient to realise the desired transformations. An important factor that also contributes to the achievement of the organisational goals is the standardisation of statistical methods and IT applications

(Braaksma 2009). This is the subject of Section 4. The availability of such standards enhances the transparency and flexibility of the design of the statistical production process, thus increasing the potential for reuse and hence reducing costs. Greater transparency in designing production processes is also expected to result in a better grip on the quality of the statistical data produced. And it is important to embed the enterprise architecture in the organisation in order to successfully implement the ensuing redesign programme. Section 5 describes two organisational arrangements that have been made: a board for architectural issues and a board for portfolio decisions. Decisions on the enterprise architecture, on standards and on the implementation of the redesign programme are to be based on cost-benefit considerations. Section 6 is devoted to this side of applying an integral approach to statistics. Section 7 summarises our experiences and highlights some issues to be dealt with in the future, including international cooperation.

2. The Need for an Integral Approach

Traditionally, statistics of most NSIs are typically produced in more or less separate and autonomous production lines, each having their own methodology and IT systems. In a nutshell, the conventional way of producing statistics consists of surveying a sample or census of persons, households, businesses or institutions with a questionnaire, processing the data using appropriate methodology, compiling tables and explanations, and publishing results on the underlying population. This way of producing statistical information places much emphasis on the development of efficient methodology for data collection, data editing and analysis. Over many years, the accepted statistical methods and techniques have become increasingly sophisticated.

The changes in the environment in which they operate have prompted NSIs to re-examine their methodological research programmes (HLG-BAS 2012). In the field of data collection, the research is now more focussed on the use of administrative registers and mixed-mode survey strategies, preferably using modes with lower per-unit cost (e.g., Internet). The new ways of data collection in turn call for advanced estimation methods for combining data from various sources, or for small areas (domains) where by design few data points are collected. The ambition to publish statistics more rapidly or work more efficiently has likewise driven improvements in macro-level editing and imputation techniques.

However, new and improved methods alone are not sufficient to completely attain the four strategic goals SN has set for itself, as they leave the traditional approach to producing statistics intact. There are two main reasons why retaining autonomous production lines may impede reaching the goals.

First, as independent production lines for statistical products typically have their own design, customised methodology and IT solutions, they are expensive to develop and maintain. Adopting a more integral approach to designing statistical production processes creates opportunities for economies of scale in process design, statistical methods, IT applications, development, training, and so on.

At SN, the costs of IT maintenance are particularly relevant. A decade ago, the statistical departments often had their own IT solutions, which were in many cases maintained locally. In fact, SN still has to deal with a large legacy of IT systems and their increasing maintenance costs. The sustainability of such solutions was eventually questioned. Incidentally, many large organisations face or have faced these problems, which can be seen as a natural corollary to the exponential growth of IT in the last few decades (Engelbertink and Vogt 2010).

Second, although the designs of autonomous production lines may be optimised from a local point of view with respect to certain dimensions such as output quality and processing costs, there is no incentive to optimise these designs from a corporate point of view. Especially issues dealing with the coherence of statistics, but also the reduction of the reporting burden and the reuse of IT solutions should be tackled companywide. Moreover, autonomous production lines make it difficult to combine and reuse already available statistical data, which is necessary for a quick response to new information needs.

Full optimisation is the Holy Grail of so-called total survey design. However, a more integrated approach to statistics production requires optimising the set of production lines as a whole, that is, across the whole organisation. Reduction of response burden, for instance, is possible when the same source data can be used in more than one production line and, conversely, a specific set of statistical tables may be based on a number of sources. This leads to an interconnected set of production lines, that is, a network, rather than the traditional set of autonomous lines. In such a situation, one should ideally aim at total *network* design rather than total *survey* design. The network is optimised for the same dimensions as is the case when applying total survey design, but the optimisation is global rather than local. This does not imply that tailor-made production lines have to be totally banned, only that they have to be subjected to an organisation-wide assessment.

The need to design production lines collectively from a company-wide perspective calls for an enterprise architecture, which serves to a link pin between the strategic goals at the level of the organisation with the individual production lines used for producing the statistics.

3. An Outline of the Enterprise Architecture for Statistics Production

When applying an integral approach to statistics production, this does not mean that everything has to be designed at the same time. Rather, one can begin by putting up a framework which gives overall guidance when successively redesigning the production processes of the organisation, in such a way that the strategic goals of the organisation are met. Such a framework is called an enterprise architecture.

In the enterprise architecture of SN, the strategic goals are translated into a future model of making statistics as well as a number of principles that underpin this model. Naturally, this future model provides for the possibility of integrated production processes. It serves as a reference framework for redesigning statistical production processes and their supporting IT systems. The reference framework is mandatory in the sense that redesigned processes have to be consistent with the enterprise architecture on the basis of ‘comply or explain’. A governance process has been implemented accordingly. This process is described in Section 5.

Within an enterprise architecture, a distinction can be made between an overall business and information architecture on the one hand, and a software and infrastructure architecture on the other hand (Van ’t Wout et al. 2010). The software and infrastructural needs follow the business and information needs. In this article we focus on the business and information architecture (Huigen et al. 2006/2009). We will broadly describe the contents of the architecture as developed by SN, emphasizing its key features. The architecture will be illustrated by an example in the domain of economic statistics, showing the added value of this approach.

3.1. Future Model of Making Statistics

The future model of producing statistics is founded on a number of key principles. We briefly mention the four most important ones.

- **Output orientation**. SN wishes to publish *relevant* statistical information of high quality, tailored to user needs. These needs are leading when (re)designing statistical production processes within budgetary constraints. This implies that the output requirements need to be specified at the earliest stages of the redesign. The specifications include quality requirements. Legal obligations, in many cases captured in European regulations, are to be met at all times. Other less formal information needs must be established and formally decided on by SN on the basis of cost-benefit considerations given the strategic goals.
- **Accessibility and comprehensiveness of the statistical information**. For users of statistical information it is important that the information is both accessible and comprehensive. This is true for external users as well as for staff within SN that make use of statistical information as input for further processing. The principle not only implies that the populations, the variables and the reference periods to which the data relate are clearly defined, but also the quality of the data. The relevant quality dimensions depend on the use to be made of the information. Such data describing information is called metadata. When (re)designing a production process, this kind of metadata needs to be considered explicitly: there is no data without metadata. Although this may perhaps seem obvious, practice shows that it is often not so easy to comply with this principle. It also contributes to the strategic goal to improve coherence.
- **Reuse of data and metadata**. The policy at SN for data collection stipulates never to ask for information by questionnaire if that information is available in any administrative register. Only if it is unavoidable can information be collected directly from respondents, but even then the information can only be asked for once. Moreover, the use and reuse of source data and intermediary results by different internal users not only supports reduction of the reporting burden, but also contributes to higher cost effectiveness. A prerequisite for reusing data and metadata is that the available statistical information is accessible to internal users (see previous principle). Another prerequisite is, of course, that there are no legal obstacles to reusing data. Fortunately, the Dutch Statistics Act actually promotes the reuse, for statistical purposes, of data by SN.
- **Modular production lines**. The future statistical production processes are designed by making use of customisable standard modules. The resulting production lines are cost efficient and can be adapted more easily and flexibly to changing information demand.

The model and the principles described above hold for *all* future statistical production processes of SN. It covers the production of both social statistics and economic statistics. But up till now, the implementation of production processes according to the enterprise architecture has focused mainly on the first three principles, as we will see in Subsection 3.3. In Section 4, the consequences of the fourth principle are discussed.

In addition to the four principles, the enterprise architecture is based on three notions that appeared to be useful in working out the future process model and underlying principles. These notions are steady state, fixed data interface and chain management. They have to do with the network aspects of making statistics, mentioned in Section 2. Statistical production processes are connected to each other when the output of one process is (re)used as input for another. For instance, data are collected only once and then reused where needed. But also processed data can be reused, and even disseminated data are used in a feedback loop when producing data on subsequent periods, or can be further processed in the context of other statistics. In this sense, the set of production processes can be seen as constituting a network. An example of such a network is given in Subsection 3.3.

In order for the output of one production process to be usable as input for another, this output has to be not only well defined, but also stable. Statistical data sets that fulfil these requirements are called **steady states**. They are well defined, that is, the populations, the variables and the reference periods to which the data relate are clear as well as the quality of the data. Thus, a steady state consists of statistical data and the corresponding metadata. The stability concerns their quality. In the current production lines, the quality of the data is often continuously improved, without clear standards. To the (internal) users there is much uncertainty about the version to be used as input for further processing, since there is always a new 'latest' version. In the future production lines this uncertainty is removed. Although a steady state may consist of different quality versions (for example relating to preliminary and final figures), each quality version is stable and designed in advance. For short, we will call each quality version a steady state.

For reuse it is important that the steady states are internally accessible. To that purpose the enterprise architecture of SN recognises a number of **fixed data interfaces**. Besides the (internal) archive function, a data interface facilitates the exchange of steady states between different production lines, including the metadata concerned. A data interface has two sides: one side to deliver the statistical data set and one side to retrieve it. The supplier of the data (supplying production line) delivers and the user of the data (demanding production line) retrieves. For the physical realisation of the data interfaces, an organisational unit called the Data Service Centre (DSC) has been created. It archives the steady states and facilitates the reuse of data and metadata.

The third notion is **chain management**. It concerns the planning and coordination of the exchange of steady states between the statistical production lines (Van Delden et al. 2009). This is distinguished from process management, which is the management within a single statistical production line with steady states as outcomes. Chain management becomes more important as the reuse of steady states grows and hence the interdependency between the statistical production lines increases, that is, when we move further away from autonomous production lines towards an integrated network.

Figure 1 gives a schematic representation of the ideal to-be situation for the production of statistics at SN. In the figure a distinction is made between the design of the production processes (upper layer), the management within and between the production processes (middle layer) and the implementation of the production processes (lower layer). The diamond shapes represent activities, while the rectangular shapes represent information products.

On the right-hand side, the design of the production processes starts by exploring the information needs. Based on these information needs the steady states including quality standards, as well as the production processes and methodology are designed, depicted as ‘design data’ and ‘design process’, respectively. Thus, the design is output oriented in accordance with the first principle, and is restrained by the strategic aims of the organisation and by policy, in particular budgetary resources. The design of the data results in the specifications of the steady states. The design of the statistical production process results in production rules for the modules mentioned in connection with the fourth principle above. This is elaborated in Section 4. The specifications of the steady states and the production rules are both called metadata.

The implementation of the statistical production process actually produces the steady states and consists of working through a series of activities from data collection through dissemination. Note that the figure does not depict the dynamic nature of the statistical production process as it is possible to go repeatedly through parts of the process by reusing data.

Data to be shared among different production lines are exchanged through fixed data interfaces, symbolized at the bottom of the figure by ‘data (steady states)’.

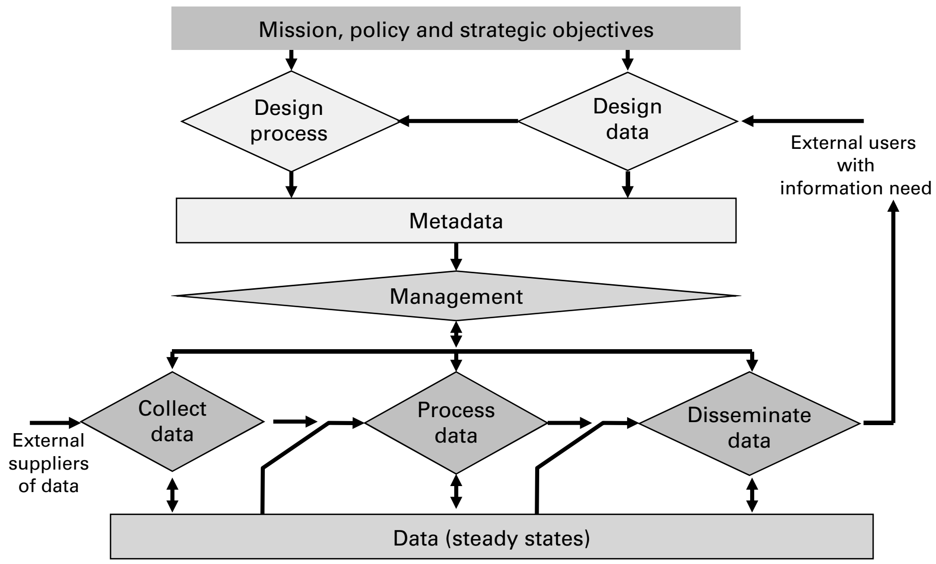

Fig. 1. (Simplified) model of future production of statistics at SN

3.2. Applying the Enterprise Architecture in Projects

At SN, redesign of statistical production processes always takes place in the form of projects. Related redesign projects are brought together under a programme in order to coordinate the dependencies between the projects and the deployment of resources. Usually the following disciplines are involved in a redesign project:

- Business analysts and project architects, who design the statistical production processes, compliant with the business and information architecture.
- Methodologists, who are responsible for sound methodology.
- Information analysts and software architects, who define the (IT-)use cases, the logical view and the implementation view of the production line to be built, compliant with the software and infrastructure architecture.
- Software developers and testers, who actually build and test the production lines.
- Statisticians, who are eventually going to use the production lines and who feed the project with subject-matter knowledge.

When applying the enterprise architecture in actual redesign projects (for an example see the next subsection), we encountered a number of practical issues. We mention the most important ones. First, the architecture reflects the corporate interest of SN. Often no local need to apply the architecture is perceived. Second, the architecture was initially developed and formulated top-down, without much explanation as to how and why it should be applied in practical situations. This led to misunderstandings. Third, there is a need for feedback to those who design the architecture from the level of application – by various disciplines that are involved in the redesigns projects – in order to get a practical set of architectural guidelines. These issues are related and show that the top-down approach has to be balanced with a bottom-up approach.

In order to overcome these issues, a number of measures were taken. One was to form multidisciplinary project teams consisting of all the disciplines described above, including explicit involvement of project architects. A second measure concerned the governance structures that were set up. This is the subject of Section 5, which describes the context in which redesign projects operate. Feedback information from both operational and strategic areas has resulted in a number of improvements in the architecture.

3.3. Applying the Enterprise Architecture: an Example

The enterprise architecture has been applied and further elaborated in a programme for the redesign of economic statistics (Braaksma 2007; Braaksma and Buiten 2012). This programme ran from 2007 to 2011. It focused on the core of statistics relating to economic growth, in particular annual business statistics and related short-term turnover statistics, which are covered by two European regulations: the Structural Business Statistics (SBS) and the Short Term Statistics (STS) regulation respectively. The system of National Accounts also fell within its scope. The programme centred on three strategic goals: increase efficiency[2], reduce administrative burden and improve quality.

[2] That is, to reduce the related staff by some 20%, from 325 fte to 260 fte.

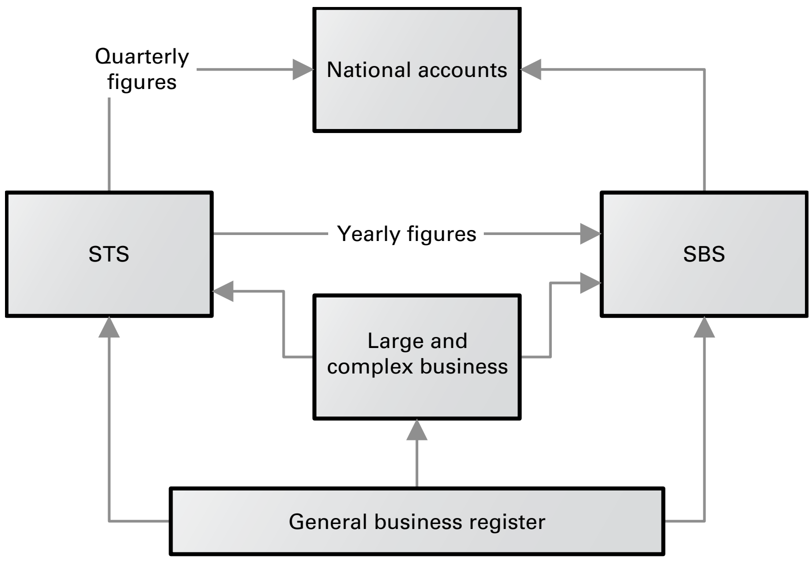

Fig. 2. Dependencies within a set of production lines for economic statistics

This redesign programme explicitly followed an integrated approach based on a comprehensive architecture for the statistics concerned. That is, the compilation of the different economic statistics is no longer designed as a set of separate production lines (stovepipes) but rather as an interconnected set of related statistical production processes. Figure 2 shows the most important statistical production processes and their interdependencies. In the following we concentrate on the STS part of the programme.

In the context of the STS regulation, SN publishes the period-to-period growth rates of monthly, quarterly and yearly turnover for four major economic domains: manufacturing industry, retail trade, construction and other services (e.g., transport). Turnover corresponds to market sales in goods and services to third parties. These high-profile short-term statistics are used to analyse and monitor current economic developments at both the national and European level.

The STS output data are growth rate indices, which provide information as to how turnover has risen or fallen compared to the previous period (month or quarter). Information on the structure of the economy for a specific year is provided by the annual production statistics based on the SBS regulation. The STS and SBS, in turn, form input for the National Accounts. For example, the short-term turnover statistics are used as a source for the compilation of quarterly Gross Domestic Product (GDP) estimates.

Before the redesign, there were coherence problems between the turnover growth figures produced by STS and those of SBS and National Accounts. The differences were especially marked for large and complex business units. The coherence problems were difficult to trace and solve, because the STS and SBS production lines were designed independently of each other. In the new design, data for the large and complex business units are collected and processed in a separate production line, generating turnover figures which STS and SBS are obliged to use. This partly solved the coherence problems. To further deal with the inconsistencies, the STS turnover figures for the smaller and less complex business units are used by SBS to calibrate its yearly figures. Furthermore, the STS figures are now mainly based on value added tax (VAT) data instead of survey data. Use of the VAT register also added to the reduction of the reporting burden, another important aim of the redesign.

Central to this way of compiling economic statistics is the General Business Register, maintained by SN. From this register, population frames are derived once a month and these are used to delineate the populations of STS, SBS and the large and complex businesses in a coordinated way. They also provide a central linkage of business units to administrative units, that is, fiscal units in the VAT register.

The interrelated production lines described above and depicted in Figure 2 were designed simultaneously in two stages. In the first stage, in line with the architectural principles, the separate products to be generated by the production lines for STS, SBS, large and complex businesses and the General Business Register were specified in detail, including quality requirements and delivery dates. This resulted in properly defined steady states relating to the different periodic growth figures and the population frames that could be exchanged according to the agreed delivery dates. In fact, different quality versions of the steady states were defined. All parties involved, including National Accounts as a major user of the turnover figures, collaborated in this stage, which can be seen as a form of chain management by matching demand and supply. In addition, agreements were made on error handling in the monthly population frame and about dealing with (too) large differences between the turnover totals of STS based on VAT data and SBS based on survey data. Note that the new design has not completely eliminated the inconsistencies between STS and SBS. This first stage corresponds to the component ‘design data’ in Figure 1.

Given these joint agreements on the data to be produced, that is the steady states, the individual production lines could be further designed relatively independently of each other in a second stage. In Figure 1, this stage is represented as ‘design process’. If data has to be exchanged between various users the production lines are ideally connected to a standard data interface. However, at the time the programme was carried out, this principle did not have the highest priority. The standard data interfaces, which are realised by the DSC (see Subsection 3.1 above), were not fully operational. Similarly, the various production lines were designed as much as possible according to the same pattern, but not modular. The approach of modular production lines is still in its infancy. In the next section we will elaborate on this approach.

4. Modular Production Lines and Standardisation

The architectural principle of using standard modules for statistical production lines may seem obvious, since the reuse of parts of production lines has a high potential for cost reduction. However, there are a number of complicating factors. First of all, the modules can be viewed either from a business (process) perspective or from an IT (system) perspective. These different viewpoints are not always recognised and may cause miscommunications. In both approaches it is difficult to choose the most useful modules. Especially for the IT modules, complications become apparent when considering the degree of parameterisation, performance requirements and integration problems between modules. Moreover, there is a limit where further standardisation is no longer efficient in terms of costs or flexibility. We consider achieving optimal standardisation a main challenge of the redesign programme of SN. We have converged to the following approach, which starts from a business perspective rather than an IT perspective.

Roughly, when designing a modular production process we find it useful to distinguish between the design of the ‘know’ and its ‘flow’. The know involves the choice and specification of the statistical methods, such as imputation methods or estimation methods. The flow involves the deployment of the know: the process modules in which the statistical methods are encapsulated, the order of these modules, the decision points that manage the statistical data through the modules, the triggers to start or end a module, and the process indicators to manage the flow as a whole. A trigger may be quality driven, cost driven, time driven or event driven.

Note that the process modules provide the link between the know and the flow. We will explain this by means of an example. Suppose we have a statistical dataset which has to be validated and corrected for so-called 1000-errors, such as made by respondents to surveys who overlook that turnover has to be given in thousands of Euros. In order to do so, a standard statistical method is to compare the data in this dataset, that have to be validated and corrected, with reference values. The discrepancy with the reference value is compared to a norm. If the discrepancy is too high, a correction is carried out. If this is not the case, the value is retained. Quality indicators are used to indicate the validation or correction of the data. So the process module has an input value, a reference value and a discrepancy norm as inputs, and an output value and a quality indicator as outputs for each data item that has to be processed. The input values are assumed to be available. A common method to obtain reference values is to take group means. The question is whether these group means should be determined within the validation module or that we should define a separate averaging module for obtaining group means; or even more generically, a module to estimate population means/totals. This example shows the difficulty of choosing the scope of an individual module and the level of granularity of the set of modules.

Although statistical methods are a key factor when constructing modules, the architecture itself does not make methodological choices. Instead, it prescribes the standard methods that may be used. Recently, the enterprise architecture of SN has been complemented by a series of standard methods. This series can be seen as a catalogue of validated and approved statistical methods, presently used at SN, that can assist in the design of a production process. By using only approved methods included in the series, the quality of the output of a production process can be better substantiated. A first version of the series was completed in 2010, and will be updated at regular intervals with new and improved methods. For example, a method to validate and correct 1000-errors is described in this series.

The process modules described above are the natural starting point to define IT modules. These IT modules do not necessarily correspond one-to-one to the process modules. Because of performance reasons, IT modules may consist of more than one process module. If process modules are commonly used together in statistical production processes, this may be taken into account when building IT modules.

An important consequence of the modular approach is that IT modules should be rule based. Standard methods, and the process modules that use them, can involve complex parameterisation (or specification). We call an IT module rule based if its control parameters can be specified by non-IT specialists (for example subject-matter specialists or methodologists) by means of an easily accessible interface, without need to change the internal code of the IT module itself. The extent to which an IT module should be rule based is an implementation choice, and touches on the issue of local and global interests. This choice depends on implementation costs, maintenance costs and the expected need to adjust the specifications. In general, the implementation costs will rise as the IT module is more rule based. Similarly, the costs for specification of rules will rise. But on the other hand, the more generic rule-based IT modules can be used in more different statistical production lines.

A key issue to consider is interoperability. When putting together different IT modules into a single statistical production line, one might encounter a problem of integration because the IT modules speak different languages and do not fit seamlessly together. This requires either a conversion effort between IT modules or a standardised exchange format ('plug and play').

In order to facilitate the (re)design of statistical production processes, one could consider developing a library of standard process modules. A number of potential candidates is currently being identified and described (Camstra and Renssen 2011). In addition to this library of standard process modules, one may also consider a library of standard IT modules. These IT modules are linked to the process modules, are rule based and are interoperable. The IT modules can be exploited to quickly and efficiently design and build a rule-based modular production line. By adjusting the rule-based specifications, it would be possible to quickly and flexibly respond to changing information demand.

To give an example, SN has recently developed an IT module for macro editing (Ossen et al. 2011). The main feature of this module is to validate statistical data at an aggregate level and if desirable to zoom in and edit the data at the micro level. This module is flexible in the sense that users can specify their own levels of aggregation for different types of units. As a policy, the degree of flexibility of an IT module should be such that it is applicable in at least two statistical production processes. This is also the minimum rate for developing a standard IT module. The idea is that other statistical production processes will follow. Naturally, this is stimulated during the redesign projects. Currently, one year after the release of the IT module for output validation, it has been successfully integrated in four statistical production processes including the redesign of economic statistics described above.

We conclude that we have not solved the standardisation challenge yet, but we have made progress. Statistical methods are standardised and reused. A library of standard process modules is under development. But we do not yet have a library of standard IT modules, nor do we have rules to balance costs, flexibility to accommodate changes in available input or required output, and other aspects that are relevant when pursuing standardisation. This means that in practice judgmental choices have to be made when redesigning statistical production processes. For this, SN has put a governance structure into place. This is the subject of the next section.

5. Governance and Organisation

When embarking on a redesign of multiple statistical production processes, several aspects are relevant from a managerial perspective:

- managing the portfolio of redesign projects.
- managing individual redesign projects.
- managing the enterprise architecture.

This implies a need for specific governance provisions. These provisions have to ensure that the projects are correctly prioritised and that they are efficiently executed in compliance with architectural requirements, which in turn have to reflect the strategic goals. They have to be tuned to the existing organisation, within budgetary constraints and taking into account the organisational culture.

In order to manage the portfolio of redesign projects and to manage the enterprise architecture as such, two central boards have been created, the Central Portfolio Board (in Dutch: CPR) and the Central Architecture Board (CAB), respectively. In the following subsections, the governing and organisational arrangements are described, including these two boards.

5.1. Portfolio Management

For the prioritisation of projects at the level of SN as a whole, the CPR has been established. It is chaired by the Chief Information Officer (CIO) of SN; all internal stakeholders are represented at strategic management level.

The CPR decides on the actual redesign programme at corporate level, by selecting which projects are to be executed from the proposals submitted by the various business units (who also have their own portfolio boards). Portfolio decisions are taken in the light of strategic goals, and on the basis of business cases that are provided with each redesign proposal. A business case is a comprehensive assessment of all current and future costs and benefits; this is further explained in the next section.

The CPR also allocates resources, in particular concerning IT services, statistical methodology, process development, and project management. The CPR does not actively pursue a multi-annual plan of controlled redesign of all statistical processes. Rather, business units provide requests that have been prioritised by their own portfolio boards.

5.2. Managing Individual Redesign Projects

Once the projects have been prioritised, they are executed following the well-known Prince2 project management method (Office of Government Commerce 2009). This method ensures that stakeholders' interests are systematically taken into account. For each project a steering committee is established, in which the following main roles are distinguished (note that a steering committee member may have several roles at once):

- Executive
- User
- Supplier.

The executive is usually the owner of the statistical production process to be redesigned. The executive chairs the steering committee and is responsible for the realisation of the project. The users have an interest in the deliverables of the project. For instance, they may be involved in carrying out the statistical production process concerned or make use of its output. The suppliers provide the resources necessary for the redesign project as described in Subsection 3.2. The Prince2 project management method provides for a solution mechanism for situations in which stakeholders disagree.

The use of Prince2 in itself does not guarantee that the results of the projects are in line with the strategic goals of SN as incorporated in the enterprise architecture. In fact, as is the case with any enterprise architecture, compliance issues have to be dealt with (Foorthuis 2012). Therefore, in addition to the use of the Prince2 method, there are a number of compliance requirements to which projects have to conform. These requirements stem from the architectural framework as described in Subsection 3.1 and take the shape of templates (BAD, MAD, SAD, refering to business, methodology and software, respectively; Hofman 2011) describing the main aspects of the redesigned statistical production process. These aspects are the output data of the production process, the data to be used as input, the methodology used, the process flow and the IT landscape. In effect, the templates provide guidance for applying the enterprise architecture in projects.

Once the redesigned production process is described according to the templates, it is reviewed. For this, a review procedure has been implemented (Foorthuis et al. 2009).

5.3. Managing Architectural Issues

Evidently, the enterprise architecture as such has to be managed. That is, formal decisions have to be taken on its contents and scope. Of course, managing the enterprise architecture also implies change management.

As mentioned in Section 3, in the enterprise architecture the strategic goals are translated into a future model of producing statistics, as well as a number of principles that underpin this model. Decision making on the architecture thus involves making explicit choices on how to interpret the strategic goals and make them operational. This includes making choices on what aspects of producing statistics to incorporate in the architecture. For instance, are the methods series or the set of standard IT tools to be included in the architecture? It also includes making choices regarding the balance between local and corporate interests. This is particularly important, since the architecture is a reference framework for the individual redesign projects. Stakeholders in the strategic goals of SN are not systematically represented in the steering committees of these projects, but their interests are served by the templates (BAD, MAD, SAD) to be used by each project.

It is important to emphasise that the enterprise architecture is a point on the horizon. It is a blueprint of the desired future situation regarding the organisation's processes and IT. As a consequence, decisions have to be taken concerning the trajectory to be followed in order to approach that point on the horizon. This gives rise to issues such as the question of what architectural principles deserve priority and the extent to which compliance with the architecture has to be enforced, in particular if services such as the DSC are not fully operational yet and the set of standard IT tools and modules is still being developed. In fact, the management of such ‘libraries of standards’ is itself an issue. And again, the choices to be made have to take into account both legitimate local and corporate interests.

The Central Architecture Board (CAB) has been created to address all architectural policy questions mentioned above. It is the board where architectural principles are prioritised and compliance issues discussed. The templates used in the redesign projects are formally approved by the CAB. It also advises on the set of standard IT tools, mentioned in Section 4. At SN, decisions on IT issues are not taken by a corporate IT board, as is done in some other countries; SN does not have such a board. The reason for this is that IT is seen as the solution to business issues. These have to be solved in accordance with the enterprise architecture, of which the IT architecture is a part. This makes the CAB an appropriate board to discuss strategic IT issues (Hofman and Van der Loo 2011).

The CAB is an advisory body of the Board of Directors and is chaired by the head of the CIO office of SN. All stakeholders, either with local or corporate interests, are represented at tactical management level. In fact, the CAB is one of the places where the tension between local and corporate interests can be addressed: those carrying the costs of complying with the architecture in individual redesign projects meet with those in charge of the realisation of strategic objectives, such as the reduction of the overall administrative burden to business or cost reduction through software standardisation. The CAB has proved to be instrumental in obtaining support for the approach to redesigning statistical production processes under architecture.

The role of cost-benefit considerations in the decision making of the CAB is discussed in the next section.

6. Cost-Benefit Considerations

As explained in the previous sections, the four strategic goals of SN are advanced by applying an enterprise architecture, promoting standardisation and implementing organisational provisions for managing redesign projects. However, the goal of cost efficiency is only reached if cost-benefit considerations are correctly taken into account when taking decisions, of course. We look at cost-benefit considerations first for the architecture, then for deciding on standardisation, and finally for project decisions on individual (re)design projects in the context of portfolio management.

The decision to start redesigning statistical production processes based on an enterprise architecture was strategic. Preparations were started in cooperation with an external partner, resulting in a so-called ICT Masterplan (Deloitte and Statistics Netherlands 2005; Ypma and Zeelenberg 2007), which envisaged an architecture for SN as a whole, the redesign of virtually all statistics under this architecture and central provisions for its implementation. The plan was not based on a detailed, quantitative, enumeration of costs and benefits through time, allowing an analysis of return on investment. Rather, preliminary studies by the external partner showed the strategic necessity to embark on the redesign of statistical production processes within an architectural framework. Basically, these studies provided the reasons for applying an integral approach, as described in Section 2. One of the supporting observations made was that most large information processing organisations have had similar experiences to those of SN: a proliferation of software tools at the end of last century, followed by a drive for standardisation.

The fact that the decision to develop and apply an enterprise architecture was based on strategic considerations and perceived logic rather than a detailed insight into costs and benefits does not mean that subsequent investment decisions cannot be based on cost-benefit analyses. On the contrary, where possible, this should be done. Once the architecture is in place, two types of investment decisions can be distinguished: the decision to develop a standard process or IT module or tool, for instance for data editing, and the decision to redesign a specific statistical production process or set of processes within a common architectural framework. These two types of decision are intertwined. The benefits of a standard IT tool or module will increase with each process that is redesigned to use it, and a redesign becomes more attractive the more standard IT tools or modules are available to support it. On the other hand, investing in IT standards has to be based on the assumption that sufficiently many redesigned processes are going to use them, and investment decisions on new statistical processes must be made under the assumption of availability of necessary IT tools and modules.

6.1. Architecture Costs and Benefits

As already said, the decision to start developing an enterprise architecture was not based on explicit and quantitative cost-benefit considerations. It is possible, however, to give some indications on both aspects in hindsight. Below we give an overview of costs that have been made and benefits that have been realised so far.

On the cost side, we can distinguish between initial investment costs, maintenance costs and support costs. The initial investment concerned a core team of two external consultants and three internal experts who worked almost full time during a six-month period in 2006 on drafting the enterprise architecture. For the business architecture in particular a lot of pioneering work had to be done, since no examples existed at the time. Many elements had to be developed from scratch. Many other people were involved part time, for example for reviewing drafts and for supplying subject-matter background knowledge.

After the initial development phase, a small team of architects was established including two lead architects (one for the business architecture, one for the IT architecture). Their task is to maintain, promote and further develop the enterprise architecture. They also prepare the meetings of the CAB. The architects combine their architectural work with other duties like consultancy, research, business analysis and software development.

Another cost element is training in architecture. A first batch of people was trained by external consultants using a standard off-the-shelf course. This was considered not sufficiently tailored to internal needs and therefore in 2009 a three-day course on business architecture was developed in-house. The internal course has since been given annually for staff who need more than a casual understanding of architecture (business analysts, methodologists, software developers). They highly appreciated the course.

Thus, the bulk of costs concerns human resources. Apart from this, some limited costs arise from software tools supporting architecture development.

On the benefit side, most rewards materialise through application of the enterprise architecture in concrete (re)design projects. Various benefits are reaped when applying the architectural approach: for example, statistical production processes run more smoothly, cost less, are better integrated into the existing environment and are more transparent, which in turn enhances quality and speed and reduces risks. It is difficult, however, to quantify such benefits since most of them are not easily measured, the role of architecture is intertwined with other aspects (like project management methods, software development methods) and we cannot conduct falsifiable or repeatable experiments. This situation is acknowledged in the scientific community (Liimatainen 2008), although some attempts are being made to measure benefits (Boucharas et al. 2010).

Larger-scale redesign programmes may reap additional benefits from an architectural approach by developing a programme architecture at an early stage. A clear example is the redesign of economic statistics introduced in Subsection 3.3, which has benefited greatly from the comprehensive programme architecture that was developed in early 2007. The programme architecture enabled integral management of the development programme, and scoping of individual projects was simplified. Moreover, a new and efficient organisational structure for economic statistics has been created along the lines of the programme architecture. And while the main role of the programme architecture was at design time, it still plays a part at run time. This is most obvious in chain management, where the concept of steady state is rather important and is used heavily to manage quality expectations and realisations.

Furthermore, an important indicator for the perceived benefits of the enterprise architecture is its acceptance by the business, in particular key players in subject-matter domains. An audit by a team containing external auditors showed general appreciation among persons interviewed (internal report by the audit team, consisting of D. Baart, A. Cholewinska, J. Jansen and Y. Vergeer, 2011). In particular, the steady state concept was highly valued. Concerns expressed relate for example to scalability – how much architecture is needed for smaller projects? – and communication. The fact that appreciation for the architecture has increased is probably due to a number of key factors. First, the business now better understands the enterprise architecture and its uses. Second, a continuous dialogue and interaction of architects with end users has helped to improve the architecture and the way it is being applied. Third, the continued support and commitment from strategic management for the architectural approach, as seen in the CAB, was crucial.

When the notion of architecture was first introduced, it met with scepticism and its added value was questioned. The scepticism has largely disappeared over the years and the added value is now broadly recognised. Architecture has become an established discipline.

6.2. Standard IT Tools and Modules

As to standard IT tools and modules, it has proved difficult to give decision making a sound financial footing, although development or purchase costs can be estimated. A well-founded judgment requires an overview of processes where a potential standard tool or module is to be applied, insight in the costs of implementation (including interfaces), an idea of training costs, a comparison of old versus new maintenance costs, life cycle estimates, and so on. As a result, decision making on IT tooling is based on qualitative rather than quantitative considerations.

Applying potential new standard IT tools or modules in a pilot setting may help to get a feeling of the costs and benefits involved. This approach is implemented as one of the key elements of the Innovation Programme that Statistics Netherlands has started in 2012 (Braaksma et al. 2012). Although the experience at the time of writing is still limited, already a number of successful pilot studies of potential standard IT tools and modules have been carried out, in close cooperation between potential users and IT staff. The fact that potential users are actively involved helps to ensure that a new tool actually satisfies business needs and hence improves the chances that it will be actually used. One new tool (Microsoft PowerPivot, an Excel add-on for processing large datasets) is, after a successful pilot study, submitted for inclusion in the set of standard IT tools mentioned in Section 4. Others are expected to follow.

6.3. Individual Redesign Projects

As mentioned in the previous section, project decisions in the context of portfolio management are based on their business case. This comprehensive assessment of all current and future costs and benefits includes financial as well as nonfinancial aspects. It includes, for instance, direct project costs, current and future maintenance costs, training costs, depreciations, required staff, effect on quality, response burden, continuity considerations and strategic consequences. Proposals for redesign are usually compared to alternative scenarios, including as a minimum the scenario of not starting any change at all. A manual for writing good business cases has been developed with the help of external experts. Business cases are assessed by controllers and other internal staff. When choosing between proposals that have a valid business case, the CPR is led by strategic considerations.

In practice, the problem of interdependent business cases between projects can be dealt with by starting with an integrated business case in which costs and benefits may not be fully quantified, but in which strategic considerations are taken into account. Individual projects may then refer to the overall business case and may assume the realisation of the other projects that are covered by the overall business case. Such constrained business cases allow for effective project assessments at their inception as well as during their execution. Moreover, the overall business case can be updated at intervals. When redesigning the chain of economic statistics, as discussed in Subsection 3.3, this was the approach taken.

As to quality, it was assumed that redesigned statistics would at least retain their previous level, in accordance with the second strategic goal as mentioned in Section 1. This assumption was not further specified, although the quality aspect of coherence was expected to improve. Actually, apart from sampling errors and response data, for many statistics most other quality dimensions are not routinely quantified. It was envisaged that applying architectural standards, such as specifying the quality dimension of datasets, would help reduce the weaknesses concerning quality information of statistics (Struijs 2005). For example, in the redesign programme of the chain of economic statistics this was taken into account by explicitly addressing quality indicators in chain management and its procedures. Unfortunately, it is too early to draw conclusions about the actual effects of the redesign of statistical production processes on quality, because results are only now emerging. The proof of the pudding will be in the eating: it is up to the users of the statistical information to assess the results of the redesign efforts.

7. Experiences and Lessons Learned

Redesigning statistical production processes within a common enterprise architecture has been underway for about seven years. What can be learned from the Dutch experiences? On the positive side, we conclude that an enterprise architecture for an NSI can be developed and applied. Although not fully mature, the architecture of SN is accepted throughout the institute as the mandatory reference for the redesign of statistics.

What has actually been achieved? SN has set up a Central Architecture Board (CAB) to address all architectural policy questions. Templates are available to guide the redesign projects. These templates are approved by the CAB and incorporate the essence of the enterprise architecture. For example, the principles of output orientation and comprehensiveness of statistical information are embodied in the templates. The notion of steady state has proved to be most useful in redesign projects, as illustrated by the example in Subsection 3.3. It enforces discipline: those responsible for statistical production processes have to explicitly specify their products and their potential customers. SN has implemented a Data Service Centre (DSC) to archive and exchange steady states, although its actual use has so far been limited to the archive function.

Of course the enterprise architecture is not an end in itself, but should be instrumental in bringing the strategic aims of the organisation closer to realisation. The question is to what extent the enterprise architecture has been successful in doing this. Due to the application of the concept of steady states, the comprehensibility of the data has been improved as well as the transparency of the production process. This has primarily advanced the goals of high quality standards and cost efficiency.

What has been achieved so far can be seen as an important first stage. We feel much more can be achieved to advance the strategic goals, in particular the cost efficiency. After all, the enterprise architecture only outlines a point on the horizon. In order to get closer to this future situation we need to further develop the DSC so as to facilitate the exchange of steady states. Furthermore, we have to invest in the libraries of process and IT modules, but there are still many issues in this area to be addressed. On the one hand, these issues concern the difficulties in identifying reusable modules and choosing the optimal degree of (rule-based) parameterisation. On the other hand, for many modules the business case is not well established. Especially for the IT modules the investment costs are considerable. In this area, there may be opportunities for international cooperation. This will alter the business case, since investment costs will be shared. But this requires considerable coordination effort.

Many institutes face the same challenges. In fact, they have started cooperating internationally. See for example the vision of the High-Level Group for strategic developments in Business Architecture in Statistics, established in 2010 by UN/ECE (HLG-BAS 2011). As a start, this vision resulted in a common language and structure regarding statistical process and information modelling. The international reference for process modelling is GSBPM and for information modelling the newly developed GSIM (Generic Statistical Information Model). A joint study conducted by Scandinavian countries and the Netherlands (Booleman and Linnerud 2010) has shown that the NSIs of these countries already have similar architectures when mapped onto GSBPM, which could be a good starting point for identifying promising cooperation efforts.

The achievements described above have not been realised in a straightforward way. We have learned a number of lessons, some positive, some negative, that deserve to be made explicit for the benefit of other NSIs. We mention the following three:

- Initially, support for an enterprise architecture was not general. Backing by the top level appeared to be essential to get over the critical point. The CAB played a crucial role in adapting and accepting the enterprise architecture. Two shortcomings in the architecture had to be corrected. First, the enterprise architecture focused too much on corporate interests and neglected the local interests. Second, the enterprise architecture was developed and formulated top-down, without much practical explanation as to how and why it should be applied.
- Complying with the enterprise architecture requires the fulfilment of certain organisational arrangements. For instance, compulsory use of standards implies the availability and accessibility of these standards. The same holds for reuse of data. As long as these organisational arrangements, like the DSC, are not fully implemented, there is much miscommunication and frustration when these standards are prescribed. In fact, the lack of the necessary organisational arrangements affects the business case of an enterprise architecture negatively.
- Although a good business case is essential for a good project, there are two caveats when applying it in the context of the redesign portfolio. First, not all costs and benefits can be expressed in financial terms, forcing a comparison of apples and oranges and causing a bias in favour of measurable aspects. Second, costs and benefits may depend on the realisation of other projects and vice versa. In such cases an integrated business case is called for, which may be awkward from a managerial point of view. In particular, standardisation benefits materialise only if adoption is sufficiently massive. This does not mean that the effort of specifying all costs and benefits does not make sense. It does, but the available information has to be used wisely.

How to proceed? Directed by the CAB and in cooperation with the business units concerned we would opt for an incremental approach. For example, we could first further develop reusable process modules and apply them in actual redesigns. Only after successful evaluation is it worthwhile to move to the next level and focus on creating reusable IT modules, starting with a proof of concept. Implementation of an IT module as a standard depends on cost-benefit considerations.

8. References

Bae, L.-C. (2010). Building an Enterprise Architecture of Statistics Korea. Paper presented at the Meeting on the Management of Statistical Information Systems (MSIS 2010), Daejeon, 2010. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2010/wp.7.e.pdf (accessed January 29, 2013).

Booleman, M. and Linnerud, J. (2010). Cooperation Based on the GSBPM. Paper presented at the Work Session on Statistical Metadata (METIS), Geneva. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.40/2010/wp.12.e.pdf (accessed January 29, 2013).

Boucharas, V., Van Steenbergen, M., Jansen, S., and Brinkkemper, S. (2010). The Contribution of Enterprise Architecture to the Achievement of Organizational Goals: Establishing the Enterprise Architecture Benefits Framework. Technical Report UU-CS-2010-014 June 2010, Department of Information and Computing Sciences, Utrecht University. Available at: http://www.cs.uu.nl/research/techreps/repo/CS-2010/2010-014.pdf (accessed January 29, 2013).

Braaksma, B. (2007). Redesign of the Chain of Economic Statistics in the Netherlands. Paper presented at the Seminar on Registers in Statistics, Helsinki. Available at: http://www.stat.fi/registerseminar/sessio3_braaksma.pdf (accessed January 29, 2013).

Braaksma, B. (2009). Redesigning a Statistical Institute: The Dutch Case. In: Proceedings of MSP2009, Workshop on Modernisation of Statistics Production 2009. Available at: http://www.pubkat.scb.se/Grupp/Produkter_Tjanster/Kurser/ModernisationWorkshop/final_papers/H_2_business_models_Braaksma_final.pdf (accessed January 29, 2013).

Braaksma, B. and Buiten, G. (2012). Redesign of the Chain of Economic Statistics in the Netherlands. Paper presented at the Fourth International Conference on Establishment Surveys, Montreal.

Braaksma, B., Colasanti, C., Falorsi, P.D., Kloek, W., Vidal, M.M., Museux, J-M., and Szep, K. (2013). Standardisation in the European Statistical System. Paper prepared for the NTTS 2013 conference. Available at: http://www.cros-portal.eu/sites/default/files/NTTS2013fullPaper_202.pdf (accessed March 1, 2013).

Braaksma, B., Verbruggen, M.H.M., Heerschap, N.M., and Roos, M.R. (2012). Innovation at Statistics Netherlands. Paper presented at the UN/ECE Seminar on New Frontiers for Statistical Data Collection, Geneva. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2012/mtg2/WP7.pdf (accessed January 29, 2013).

Camstra, A. and Renssen, R. (2011). Standard Process Steps Based on Standard Methods as Part of the Business Architecture. Paper presented at the 2011 ISI Conference, Dublin. Available at: http://isi2011.congressplanner.eu/pdfs/650238.pdf (accessed January 29, 2013).

Deloitte and Statistics Netherlands (2005). Excerpt from the ICT Masterplan "Counting on Statistics". Report prepared for the Advisory Council on Methodology and Information Technology.

Doherty, K. (2010). How Business Architecture Renewal is Changing IT at Statistics Canada. Paper presented at the Meeting on the Management of Statistical Information Systems (MSIS 2010), Daejeon, 2010. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2010/wp.2.e.pdf (accessed January 29, 2013).

Engelbertink, F.P. and Vogt, H.H. (2010). How to Save on Software Maintenance Costs. White Paper, Omnext BV, Veenendaal. Available at: http://www.omnext.net/downloads/Whitepaper_Omnext.pdf (accessed January 29, 2013).

Falorsi, P.D., Barcaroli, G., Fasano, A., and Mignolli, N. (2013). A Business Architecture Framework for Industrialisation and Standardisation in a Modern National Statistical

Institute. Paper prepared for the NTTS 2013 conference. Available at: http://www.crosportal.eu/sites/default/files/NTTS2013fullPaper_218.pdf (accessed March 1, 2013).

Foorthuis, R.M. (2012). Project Compliance with Enterprise Architecture. Doctoral dissertation (PhD thesis), Utrecht University, Department of Information and Computing Sciences, Center for Organization and Information. Available at: http://igitur-archive.library.uu.nl/dissertations/2012-1008-200504/foorthuis.pdf (accessed January 29, 2013).

Foorthuis, R.M., Hofman, F., Brinkkemper, S., and Bos, R. (2009). Assessing Business and IT Projects on Compliance with Enterprise Architecture. In: Proceedings of GRCIS 2009, CAiSE Workshop on Governance, Risk and Compliance of Information Systems. Available at: http://tunguska.home.xs4all.nl/Publications/Docs/Assessing%20Business%20and%20IT%20Projects%20on%20Compliance%20with%20EA%20-%20Foorthuis%20et%20al.pdf (accessed January 29, 2013).

Galvin, V. and Rhodes, N. (2010). Using the Idea of Statistical Architecture at Statistics New Zealand. Paper presented at the 12th East Asian Statistical Conference, Tokyo, 2008. Available at: http://www.stat.go.jp/english/info/meetings/eastasia/pdf/t2nzpa.pdf (accessed January 29, 2013).

HLG-BAS (2011). Strategic Vision of the High-Level Group for Strategic Developments in Business Architecture in Statistics. UN/ECE. Available at: http://www1.unece.org/stat/platform/display/hlgbas/Strategic+vision+of+the+High-level+group+for+strategic+developments+in+business+architecture+in+statistics (accessed January 29, 2013).

HLG-BAS (2012). Strategy to Implement the Vision of the High-Level Group for Strategic Developments in Business Architecture in Statistics. UN/ECE. Available at: http://www1.unece.org/stat/platform/download/attachments/65962687/10+-+HLGBAS+Strategy.pdf?version=1&modificationDate=1336403439393 (accessed January 29, 2013).

Hofman, F. (2011). Redesign Approach at Statistics Netherlands. Paper presented at the 2011 International Methodology Symposium, Statistics Canada, Ottawa Canada.

Hofman, F. and Van der Loo, M. (2011). Standardisation and Differentation in Redesigns. Statistics Netherlands. Available at: http://www.cros-portal.eu/sites/default/files//Standardisation%20and%20Differentiation%20in%20Redesigns%20v1.0p3.doc (accessed January 29, 2013).

Huigen, R., Bredero, R., Dekker, W., and Renssen, R. (2006/2009). Statistics Netherlands Architecture; Business and Information Model. Discussion Paper 09018, Statistics Netherlands, The Hague. Available at: http://www.cbs.nl/NR/rdonlyres/A6F67DFB-859A-4F3E-83A0-ED8055659D4E/0/200918x10pub.pdf (accessed January 29, 2013).

Lankhorst, M. (2009). Enterprise Architecture at Work: Modelling, Communication and Analysis. Berlin: Springer Verlag.

Liimatainen, K. (2008). Evaluating Benefits of Government Enterprise Architecture. Paper presented at the 31st Information Systems Research Seminar in Scandinavia, Åre. Available at: http://www.iris31.se/papers/IRIS31-059.pdf (accessed January 29, 2013).

Lorenc, B., Fossan, Y., Holm, E., and Jansson, I. (2011). Towards an Enterprise Architecture for an NSI. Available at: http://www.cros-portal.eu/sites/default/files/S4P2.pdf (accessed January 29, 2013).

Office of Government Commerce (2009). Directing Successful Projects with PRINCE2™ London, UK: The Stationery Office.

Ossen, S.J.L., Hacking, W.J.G., Meijers, R.J., and Kruiskamp, P.H. (2011). MacroView: a Generic Software Package for Developing Macro-Editing Tools. Paper presented at the Work Session on Statistical Data Editing, Ljubljana. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2011/wp.14.e.pdf (accessed January 29, 2013).

Struijs, P. (2005). Improving the Quality of Statistics through the Application of Process Methodology. Paper prepared for the Advisory Council on Methodology and Information Technology.

Studman, B. (2010). A Collaborative Development Approach to Agile Statistical Processing Architecture – Australian Bureau of Statistics (ABS) Experience and Aspirations. Paper presented at the Meeting on the Management of Statistical Information Systems (MSIS 2010), Daejeon. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2010/wp.3.e.pdf (accessed January 29, 2013).

Swiss Federal Statistical Office (2007). Conceptual Architecture of the Statistical Information System at the Swiss Federal Statistical Office. Available at: http://unstats.un.org/unsd/dnss/docViewer.aspx?docID=1564 (accessed January 29, 2013).

Vale, S. (ed.) (2009). Generic Statistical Business Process Model, Version 4.0. Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS), Eurostat, Luxembourg. Available at: http://www1.unece.org/stat/platform/download/attachments/8683538/GSBPM+Final.pdf?version=1 (accessed January 29, 2013).

Van Delden, A., Lammertsma, A., and Van de Ven, P. (2009). Chain Management in Statistics: Best Practices. Statistics Netherlands. Available at: http://www.cbs.nl/NR/rdonlyres/F757EED3-7B6B-4C43-B8E2-313C3E2E4B93/0/200943x10pub.pdf (accessed January 29, 2013).

Van der Veen, G. (2007). Changing Statistics Netherlands: Driving Forces for Changing Dutch Statistics. Paper presented at the Seminar on the Evolution of National Statistical Systems, New York, February 2008. Available at: http://unstats.un.org/unsd/dnss/docviewer.aspx?docID=1572 (accessed January 29, 2013).

Van 't Wout, J., Waage, M., Hartman, H., Stahlecker, M., and Hofman, H. (2010). The Integrated Architecture Framework Explained. Berlin: Springer Verlag.

Ypma, W.F.H. and Zeelenberg C. (2007). Counting on Statistics; Statistics Netherlands' Modernization Program. Paper presented at the UnStatCom/ECE Conference of European Statisticians, Geneva. Available at: http://www.cbs.nl/NR/rdonlyres/61D36740-0489-4A76-B243-E86A3BEFBBD8/0/200932x10pub.pdf (accessed January 29, 2013).

Received December 2011
Revised April 2012
Accepted November 2012

Journal of Official Statistics, Vol. 29, No. 1, 2013, pp. 73–97

Innovative Production Systems at Statistics New Zealand: Overcoming the Design and Build Bottleneck

Allyson Seyb[1], *Ron McKenzie*[1], *and Andrew Skerrett*[1]

In 2011, Statistics NZ embarked on an ambitious ten-year programme of change called *Statistics 2020 Te Kāpehu Whetū: Achieving the statistics system of the future.* This article outlines a key component of this approach to transforming how Statistics NZ delivers its statistics. This involves modernising all aspects of statistical production – from identifying need, right through to dissemination – to ensure the organisation has the functionality and capacity to deliver now and into the future. Standardisation of processes, methods, tools, and systems to increase flexibility and efficiency is a key component of this transformation.

Key words: Statistical production process; business architecture; administrative records.

1. Introduction

Statistics New Zealand (Statistics NZ) is the major producer of official statistics in New Zealand and leader of the New Zealand Official Statistics System (OSS). Like other national statistics offices, Statistics NZ faces many challenges in supplying the broad range of information that is required by modern users of official statistics. The organisation strives to maintain the quality of its statistics in terms of relevance, accuracy, timeliness, accessibility, consistency, and interpretability. At the same time, the organisation is in a tight fiscal environment that requires us to minimise costs and maximise effectiveness. Statistics NZ's aim is to produce statistics that are fit for purpose in a cost-effective and sustainable way.

In 2011 Statistics NZ embarked on an ambitious ten-year programme of change called *Statistics 2020 Te Kāpehu Whetū: Achieving the statistics system of the future* (Stats 2020 in short). The Stats 2020 programme of change is extensive. It involves changes to the way the organisation works; who it works with and how; to the systems, tools, and processes used; and to the skills, attitudes, and behaviours needed for success. These elements of change are viewed from four perspectives: transforming how statistics are delivered; leading the New Zealand OSS; obtaining more value from official statistics; and creating a responsive and sustainable organisation (Statistics NZ 2010).

This article outlines a key component of the approach adopted for the first of these: transforming how Statistics NZ delivers its statistics. This goal focuses on Statistics NZ's

[1] Statistics New Zealand, The Boulevard, Harbour Quays, PO Box 2922, Wellington 6140, New Zealand. Emails: allyson.seyb@stats.govt.nz, ron.mckenzie@stats.govt.nz, and andrew.skerrett@stats.govt.nz

Disclaimer: The opinions, findings, recommendations, and conclusions expressed in this article are those of the authors. They do not necessarily represent those of Statistics NZ, which takes no responsibility for any omissions or errors in the information contained here.

role as a producer of official statistics and involves modernising all aspects of delivery from identifying need, right through to dissemination, by ensuring that it has the functionality and capacity to deliver now and into the future. Standardisation of processes, methods, tools, and systems to increase flexibility and efficiency is a key component of this transformation.

The following objectives shaped this transformation project:

- Modernising information technology (IT) systems to leverage the opportunities that new technologies offer for working more effectively.
- Reducing risk by replacing legacy software.
- Standardising processes so that staff can move between outputs without having to learn new tools and systems. This will increase efficiency and remove a barrier to innovation.
- Updating and standardising methodologies to eliminate unnecessary diversity and duplication conceptually, in terms of business and information concepts, and practically, in terms of methods and technology.
- Creating statistical methods and systems that support a growing need for more detailed data analysis.
- Maximising the use of administrative data to reduce respondent load, minimise costs of data collection, and maximise data re-use.
- Developing and applying a quality framework and quality indicators to inform users about the quality of the data before and after statistical processing and to monitor and optimise the performance of the IT system.
- Maintaining the quality of the statistics produced by ensuring they continue to meet user needs.

While Statistics NZ is only one year into the official change programme, many of the initiatives had already been established in the organisation. These included creating common processing platforms for clusters of social and business collections, using standard statistical tools, and adopting a generic Business Process Model and Quality Management programme. Stats 2020 incorporates these existing initiatives into a cohesive, recognised, and measured programme, raising both the organisation's productivity and its flexibility. A key element of this is the replacement of legacy IT systems with fewer systems that are more flexible and more standardised, so that risks are eliminated and systems are cheaper and easier to maintain and manage.

The programme of change also builds on previous work to confirm the approach to data collection that will support future information needs. The existing structure of collections at Statistics NZ is based on a traditional model of sample surveys supported by a range of administrative data sources. By 2020, the organisation's aim is that administrative data will be the primary source of information, supplemented where necessary by direct collection. Since 2008, the organisation has focused on designing statistical architectures for both business and social statistics (see e.g., Bycroft 2011; McKenzie 2008). A statistical architecture describes at a high level how future collections of survey and administrative data will be managed and combined to produce statistical outputs. These architectures have helped structure thinking and guide the development of an integrated statistical system to meet the goals set out in Stats 2020. The new IT platforms for processing household and economic collections are designed to support the new statistical architectures and Stats 2020 goals.

The environment that Statistics NZ operates in has changed over the last ten years. Societal and economic changes as well as greater diversity have heightened demand for a broader range of statistics that will support more complex analysis at greater levels of detail (Glaude 2008). Users of official statistics want statistics to be more relevant to their needs. The role and contribution of Māori (the indigenous people of New Zealand) – economically, socially, and demographically – to the growth of the country is increasing significantly. Given this, there is an increased emphasis on information for and about Māori to support decision making and understanding.

Statistics NZ's role as the leader of the decentralised New Zealand OSS, as well as a producer of official statistics, has been reconfirmed and strengthened by the New Zealand government (Statistics NZ 2010). The organisation is mandated to coordinate statistical activity across government by driving the overall performance of the OSS and ensuring New Zealand both gets the information it needs and that this is value for money, at the lowest possible cost to government, the community, and suppliers of the data (State Services Commission 2011). This places greater expectations on the organisation to deliver a more cost-effective and productive statistical system. At the same time, the fiscal environment is becoming more and more challenging. Despite the growing costs of data collection and processing, perceptions of increasing response burden, and smaller budgets, the public sector is trying to increase and improve the services they offer.

Users of official statistics expect the national statistics office to improve the quality, quantity, and "time to market" of their statistics. The availability of vast stores of administrative data across government sectors, and the ability to easily store and manipulate the data, presents a wealth of opportunities to identify and reduce duplication, increase utilisation, and improve collection effectiveness.

Addressing these drivers has led to the developments described in this article. Section 2 describes the business statistical architecture that is helping Statistics NZ to achieve the desired changes to its systems and direction. The focus of the article from this point onwards is on economic statistics and the new economic platform. Section 3 focuses on statistical development, particularly the change in IT strategy that has seen statistical development move from being an expensive drawn-out process to a process that often does not involve IT staff at all. Standardisation and automation of statistical processing are discussed in Section 4. Section 5 focuses on governance. Section 6 discusses lessons learnt and the transformation in the culture of the organisation as a result of the changes in the way statistics are produced. Section 7 has some concluding remarks.

Throughout the article, the term "subject-matter specialist" refers to analysts who process and analyse particular surveys, such as the survey of Retail Trade, and the term "methodologist" refers to analysts who work in the methodology area of the organisation. Traditionally the roles of subject-matter specialists and methodologists were quite different from each other – methodologists primarily focussed on designing surveys and subject-matter specialists focussed on processing, analysing, and disseminating the data collected. More recently within Statistics NZ, the roles have begun to overlap – subject-matter specialists are developing skills in survey design and methodologists are taking a more active role in understanding the entire statistical production process (including the technology solutions) and leading the cost-benefit analyses and decision making. When the term "analyst" is used in this article, we are referring to either a subject-matter specialist or methodologist, or both.

2. Statistical Architecture

A key objective of Stats 2020 is to make maximum use of administrative data while continuing to produce economic statistics that are fit for use. Further, the aim is for administrative data to be the first source of data for statistical outputs (Statistics NZ 2010). This key principle of the new statistical architecture has been a driver for developing the new economic statistics platform.

2.1. Background

Statistics NZ has had access to a variety of administrative datasets for many years. A tax-based business register was developed at the end of the 1980s after a value added tax (called "goods and services tax", or GST) was introduced in New Zealand. Tax data was introduced into the quarterly and annual industry estimates of financial statistics by using tax data for strata containing the smallest units. The risks to quality from using tax data were mitigated by ensuring that the tax strata contributed less than 15% of the estimate.

A growing range of user needs created pressure for the development of a new approach. The existing survey-based method provided the information needed for the compilation of estimates of GDP, but could not meet the growing demand for detailed statistical information to support policy development and monitoring. For example:

- studies of business development need to focus on small businesses, where sample coverage is limited.
- longitudinal analysis needs to be able to track the performance of businesses over time.
- the non-profit satellite account of the national accounts needs information about units with relatively small financial flows and numbers of employees, which were not well represented in sample surveys.
- existing sample surveys could not provide the detailed information needed to estimate regional GDP.
- increased use of microdata analysis increased the demand for full-coverage unit record data.

To respond to these emerging needs without increasing respondent load, Statistics NZ embarked on a strategy that places a greater dependence on administrative data. A new high-level statistical architecture for economic statistics was formally approved during 2006 and a programme of methodological development, assessment, and implementation began. The architecture for economic statistics has now been subsumed within the broader Stats 2020 programme.

2.2. Principles

The new statistical architecture for economic statistics is shaped by the following guiding principles (McKenzie 2008):

1. Information should only be collected if a clear user need has been established.
2. Information should only be collected once.
3. Administrative data will be used as the primary source of data wherever possible.
4. Surveys will only be used to fill the gaps that cannot be met from administrative sources.
5. Survey and administrative data will be integrated using a comprehensive business register.
6. Information should only be collected from units that can readily provide reliable and meaningful estimates.
7. Large complex business units will be closely managed to facilitate the collection of all the data that is needed from them.
8. Information quality will continue to be fit for purpose.
9. Reliance on administrative data will increase in incremental steps, beginning with the parts of the population for which administrative data is robust and then expanding into more difficult areas as data issues are resolved.

These principles represent a significant change in the collection and production of business statistics (Kent 2011). Using administrative data wherever possible, with surveys filling gaps in information needs, is a reversal of the previous survey-based strategy. The developments currently underway are applying an approach that combines the use of survey data for large complex businesses with tax data from the rest of the population. Tax data is used directly for most of the population and surveys are targeted at businesses or data items that make a significant contribution to statistical outputs.

Statistics NZ has a long history of using administrative data, for example using customs data to create estimates of changes in overseas trade volumes and building consents data from local councils to create a sampling frame for the quarterly survey of building activity. However, most of the administrative data currently incorporated into business statistics has been tax data sourced from the New Zealand Inland Revenue Department. Tax data is relatively straightforward to use, because the reference numbers used by the administrative agencies are already mapped on the business register, so matching and coverage issues are easy to resolve. Furthermore, the data items are financial variables that are well defined. Limited use is being made of electronic card transaction data. In the future, this will be developed further and other data sources such as livestock records and store scanner data will be explored for use in our agriculture production survey and price indices respectively, and methods to make use of these new data sources will be introduced.

Developing a platform that can process administrative and survey data meant that a number of issues specific to administrative data had to be addressed:

- Administrative data comes in huge volumes that need to be processed and stored. An extract, transform, and load (ETL) workflow tool used can load large volumes of data quickly and efficiently.
- The processes for receiving administrative data are now well controlled. Duplicate records are often received. The administrative agency updates or corrects records without any warning. Methods for versioning data, eliminating duplicates, and managing revisions to data have had to be developed.

- The large volumes of data make traditional editing methods impractical. It is not possible for a statistical processer to examine every record. The development of automated editing and correction processes has been essential for the use of administrative data. An associated benefit is that the platform can apply the same automated processes to survey data.

2.3. *Range of Methods*

The Stats 2020 objective of maximising the use of administrative data means that statistics will be produced in different ways. The range of methods for using administrative data will change and expand as new data sources are introduced and assessed and new methods are developed and tested. The following are some of the ways administrative data is used currently or will be used in the future:

- A tax-based business register will continue to be used as the sampling frame from which survey samples are derived. For example, some information about business operations cannot be obtained from administrative sources.
- Financial collections will continue to use a mix of different methods for different parts of the population.
- When the administrative data definitions are closely aligned with the concepts required for the relevant statistical output and reporting is timely, administrative records will be used directly. This approach supports the production of aggregate estimates and provides additional unit-level data for microdata analysis.
- If the concept collected in a tax form does not exactly match the concept for a particular statistical output, some data modelling might be needed to produce robust estimates. In some cases, the statistical model can be estimated using data from other units in the administrative dataset. An example of this is combining data from two-monthly and monthly GST returns to produce quarterly statistics. In other situations, sample survey data might be needed for calibration of the model.
- Data records from several tax forms might be combined with survey data to produce the required statistical outputs. For example, Statistics NZ is planning to develop a measure of quarterly operating surplus as input to an income measure of quarterly gross domestic product. To minimise respondent load, estimates for small and medium-sized enterprises will need to be derived from two alternative sources: income and expenditure from their GST return and wages and salaries from a monthly return used to collect taxes on wages and salaries at source.
- Administrative data records will continue to be used in imputation for nonresponse in surveys.
- Administrative data will continue to be used as auxiliary data to optimise estimates from surveys (e.g., generalised regression estimates).

To support future uses of administrative and survey data, Statistics NZ needs systems that can store data in such a way that allows it to be re-used in a wide range of formats and for different purposes.

2.4. New IT Platform to Support the New Statistical Architecture

Our existing economic survey infrastructure had the following characteristics that made the new statistical architecture harder to support:

- Each survey had a separate stand-alone custom-made system.
- Survey-based systems were designed around small data volumes and could not easily be extended to handle the vast volumes of administrative data.
- Linking data from different sources was arduous.
- Data structures were inflexible, making it difficult to introduce new data sources. Data was stored in highly normalised, fixed tables optimised for transaction processing. The column names in these tables were often subject-matter specific and thus rigidly defined for a specific topic. The links between tables were also fixed and again were rigidly subject-matter specific. Reporting and analysis were customised to each production system. This allows fast processing, but changes to processes are complicated.

The microeconomic platform (previously known as BESt) developed a new storage system with entirely different data architecture. This platform has introduced two major innovations that are critical to the implementation of the new statistical architecture: firstly, it is designed for analysis, and secondly, it has flexible data storage.

2.4.1. Designed for Analysis

Our existing information systems were optimised for fast transactional processing using what is known as the online transactional processing (OLTP) model for storing data. This system design maximises processing speed and minimises storage with a high degree of normalisation. A strong focus on the data for the latest survey period meant that recording and understanding the events and decisions affecting data for earlier periods tended to be secondary. An emphasis on the most up-to-date value of each datapoint meant that when correcting data, the values replaced are either overwritten or made inaccessible in a log or separate history table. Analytical capability is often achieved by exporting data to a separate dataset which is structured in a way that supports analysis by tools outside the processing system.

When developing information systems, there is always a trade-off between speed, flexibility, and analytical capability. Existing survey systems were designed to favour speed over flexibility and analysis. Speed was important in a traditional survey processing environment, when large volumes of survey data had to be entered and edited by survey operators, but loss of history is a constraint in an environment where the emphasis is on analysis.

In developing the new platform, it was recognised that Statistics NZ is shifting from its narrower focus on data collection to becoming an organisation that provides a fuller service by collecting, analysing, and disseminating information. Data processing is becoming less important, while analytical power and flexibility in dissemination are increasingly important. To support this change in direction, the new economic statistics platform has been optimised for analysis, with analytical capability being built into the basic database design.

An online analytical processing (OLAP) methodology (Thomsen 2002) has been adopted to support this change in direction. OLAP is very common in the corporate world and many national statistics offices use OLAP as a read-only output data system that is exported from OLTP survey systems. The new production systems avoid OLTP altogether and process and store data within an OLAP analytical schema.

The software change that makes this approach possible is the ability to momentarily "pivot" a "narrow" single-cell table into a "wide" table, which allows the system to imitate an OLTP system when processing data. The wide table has variables as columns and units as rows, which is the format required by tools such as Banff and many other data-processing tools. The data is pivoted from the wide table back into the narrow table required for OLAP when bulk data processing is complete.

This OLAP approach provides a time-intelligent, version-controlled, metadata-rich datastore. The increase in computer processing power means that losses in processing speed are manageable. Large volumes of data (with versioning) are accessible to analysts without significant loss of speed in bulk data processing.

2.4.2. Flexible Data Storage

The microeconomic platform datastores are also designed for flexibility. In a traditional OLTP system, the data tables have column names that match the input and output data. For example, if a survey has a question about income, there will be a column called "Income". This hardwiring of column names requires systems to be rebuilt to handle new surveys. Systems with hardwired columns are difficult to change when new data arrives. With large volumes of administrative data to process and statistics combining various survey and administrative data sources, this approach is no longer practical.

Statistics NZ has adopted a cell-based approach to data storage that eliminates the need for fixed column names. This means that building a stand-alone system for each new survey or administrative series is no longer necessary. The cellular data warehouse schema adopted by Statistics NZ takes OLAP to the extreme by adopting a single-cell table structure. In this model, each row in the narrow table contains one data item, called a fact. The other columns contain standard metadata items, such as description, frequency, and date. This ensures that all datasets have standard metadata including a record of the rules used to change the data. When data is processed, it is pivoted on the fly so that each variable becomes a column in a wide format table. When all the transactions are complete, the data is unpivoted back into the cell-based format of a narrow single-cell table. This process is fully automated.

The cellular approach has significant advantages for the production and analysis of statistics:

- Reports, queries, and processes developed for one collection work easily for another.
- Reports, queries, and processes developed for survey data can be applied to administrative data, and vice versa.
- A creation date on each row makes the data suitable for both point-in-time and longitudinal updating and analysis.
- Metadata is stored alongside data.
- The system is resilient to any change in the structure of the data.

- Administrative data and survey data can be processed and stored on the same platform.
- Metadata for survey and administrative data is stored in a way that facilitates data integration.
- Sharing data between teams is a matter of security settings allowing access to data, rather than an exercise in IT support and redesign.
- Subject-matter specialists can design and build production systems for new outputs, and maintain existing production systems, without being constrained by data storage or system barriers.
- Administrative and survey data are stored in exactly the same way, so solutions that combine survey and administrative data are easy to implement.
- The ETL Workflow can extract and load large volumes of data very quickly.

The disadvantages of this cellular format are:

- Processing performance will always be slower than a wide-column solution on the same hardware, but this becomes less of an issue as hardware gets cheaper and database technologies become more advanced.
- The cellular data structure presents new challenges and requires new skills from in-house IT developers as well as support in the form of expert advice and training from external expert data warehouse consultants.

A number of economy-wide administrative data series and surveys with simpler data structures have been migrated to the new platform. It now contains GST data and detailed financial data covering every business in New Zealand for several years, plus import and export transactions in detail for the whole economy dating back ten years, as well as quarterly surveys of manufacturers and wholesalers. The current holding is around four billion data points. Under development are the more complex collections, including the economy-wide annual enterprise survey, the annual agricultural production survey, and the business register. These latter systems demand more advanced statistical methods and require more complex integration of survey and administrative data.

Another key development currently underway is the creation of a prototype Integrated Data Infrastructure (IDI), an integrated dataset with person-level data from multiple sources. The IDI contains longitudinal microdata about individuals, households and firms.

The IDI is not a population register, but is able to be used to answer research questions about particular populations that are well represented in the linked dataset, for example students. By bringing together many datasets into one infrastructure, the IDI allows for longitudinal analysis across education, employment, migration, welfare, and business, with potential for further datasets to be added in future. This will help Statistics NZ to meet demand for more or new information, and allow more efficient responses to changes in existing data sources.

The prototype IDI is used as a research database by both government and academic researchers. Statistics NZ's standard security measures and protocols govern the management of the data used in the IDI. Statistics NZ is required to comply with the confidentiality provisions of the Statistics Act 1975 and also protocols for security in the government sector (Statistics NZ 2012). Other security measures for the IDI include the following arrangements:

- All data collections and associated electronic workspaces will be secured (access will only be authorised for project personnel who need to access data for specific tasks, and to selected IT administrators who are required to maintain the IT system).
- Datasets will not be available to third parties.
- Regular audits of individuals able to access the dataset will be made.

Statistics NZ has developed microdata access policies and confidentiality standards for the IDI that apply to potential research outputs to protect the confidentiality of information providers. These rules also take into consideration the existing confidentiality rules for each contributing dataset (Statistics NZ 2012).

3. Statistical Development

Creating statistics is an intellectual, service-based activity, akin to practising law or engineering. Just as an engineer's bridge design is not the product of a computer-aided drawing system, a statistician's time series is not a product of the computer that calculated it. Productivity in our industry is a function of the efficiency and ability of the technical statistical staff involved. Systems can complement, but are not a substitute for, this intellectual and human productivity.

With the goal of making the subject-matter specialist role easier and more efficient, the following principles guided the design of the new economic platform:

- Subject-matter specialists must have direct control of all business processes and be able to make changes easily. This is explored in more depth in Subsection 3.1 below.
- Systems must support the experimentation and creativity of subject-matter specialists.
- Repetitive tasks should be automated.
- All data must be stored so that it can be accessed and integrated easily.
- Whenever business processes and systems interfere with analyst productivity, they should be reviewed and considered for change. For example, most existing processing systems could not produce graphs that meet Statistics NZ publication standards, so data was sent to a specialised publication team to create the required graphs. This process typically took about three working days. The solution is a processing platform that can produce graphs that meet publication standards as part of the usual quality monitoring.
- The platform must support continuous improvement and allow for timely monitoring of the quality of processes and statistical outputs.

The microeconomic platform already incorporates a module for calculating sample errors. If a new measure is developed for any other component of total survey error, a generic module to calculate and monitor it can be easily added to the platform. These measurable components of total survey error can be displayed along with other quality indicators in a dashboard, which is available for methodologists when optimising a collection design and for subject-matter specialists monitoring the ongoing performance of the statistical production processes.

Data processing error is a significant component of total survey error in the quality models described by Groves and Lyberg and by Biemer (Groves and Lyberg 2010; Biemer 2010). It includes errors introduced at all stages of processing, for example during data editing and coding or during computation of weights and estimates. The versioning of data maintained by the platform allows the pre- and post-processing versions of a dataset to be compared and analysed to assess the significance of processing error.

3.1. The Design and Build Bottleneck

The generic Business Process Model (gBPM) was originally developed in New Zealand (Savage 2008) and is reflected in the Generic Statistical Business Process Model (Metis Steering Group 2012). It has, with subsequent modifications, been widely adopted by other statistics organisations.

The gBPM includes designing and building statistical outputs as a routine part of business activity. However, in practice, under the old approach a complex one-off development process and separate project was required to make substantive process changes in a given system.

"Design" and "build" have proved to be the primary performance weaknesses in the statistical supply chain. It can take years in a national statistics office to move from identification of need to the production of a new output. Large multidisciplinary development teams lead to high costs, a lengthy justification and approval process, and long development times. The cost and long lead time makes the organisation unresponsive to changing needs. Some needs are never met as they fall below cost/benefit thresholds. We call this problem the "Design and Build Bottleneck."

To resolve this problem, the new platform has adopted an "analyst-driven" approach. In this analyst-driven model, creating new or improved statistical output becomes a routine activity for subject-matter specialists and methodologists. The role of the IT department is limited to building a generic platform that can be used by analysts for designing and building statistics as well as collecting, processing, analysing, and disseminating statistics. Incorporating analyst-driven design, build, and test functionality into the microeconomic platform allows new statistical products to be developed more quickly and at significantly lower cost.

In this model, subject-matter specialists and methodologists control all the business processes and methods. Examples include microdata editing, outlier detection, imputation, adjustment of weights, estimation of model parameters, and computation of standard errors. Common methods such as historic imputation are encapsulated in standard tools (in this case Banff, a data editing and imputation system developed by Statistics Canada (Statistics Canada 2005)), which are then combined into an end-to-end production process called a configuration. A configuration, or a group of processes that will be applied to a

Fig. 1. The generic Business Process Model

particular set of data, can be changed by the subject-matter specialist or methodologist in two main ways: firstly, the parameter settings that control a process in the configuration can be altered to improve the performance of the process, for example changing the minimum cell size criteria in an imputation cell; secondly, a process can be added or removed from the configuration, for example changing an imputation method from mean imputation to historic imputation.

This model offers subject-matter specialists and methodologists a great deal of flexibility when it comes to the design of collections and processes. This flexibility significantly affects the design of products, as it is no longer necessary to get the design absolutely right prior to the development phase. This is because it is easy to optimise the design once a configuration has been created. This flexibility also affects the production process itself. The ease with which configurations can be altered means that analysts are able to alter or replace processes as needed, with the constraint that there is a limited range of standard methods available for each process (e.g., analysts can choose between the imputation methods available in Banff). It also means that strict controls are in place to differentiate between configurations used in production and those under development. Only a very few staff are given the authority to make changes to configurations that are used in producing published products, and clear guidelines and procedures are in place to control changes. Also, for a tool to become available on the platform, it must first be validated as a standard tool by methodologists and be available in the statistical toolbox.

As outlined in Subsection 2.4.2 above, flexible and consistent data storage are critical as they give subject-matter specialists direct control over both existing and new data. Below we describe four features of the microeconomic platform that support subject-matter specialists to maintain the relevance of existing outputs and to design and build new outputs quickly and easily.

3.1.1. Business Register History

Statistics NZ maintains a comprehensive tax-based business register that is used for selecting samples and integrating and classifying administrative data. The business register records industry and institutional sector classifications for every economically significant unit in the economy. Classifications on the business register are derived from multiple sources: from administrative registers for simple units and via register maintenance surveys for larger and/or more complex units. Inconsistencies between sources are resolved by applying business rules during the register update process.

Before administrative data is used in the production of statistics, each administrative unit on the dataset is linked to the business register, so the classifications from the business register can be assigned to it. This ensures that all economic data is classified consistently by eliminating the need to rely on the classifications that may have been received from the administrative system.

The microeconomic platform stores point-in-time snapshots of the business register. Samples are selected from these snapshots and coordinated so that, for example, quarterly surveys are selected at the same time on the same basis. Immediate access to complete business register information facilitates the design and development of new statistical collections and outputs.

3.1.2. Configuration Store

The configuration store is an important innovation, because it makes the ETL workflows described in Subsection 2.4.2 user configurable. The configuration for a statistical output lists all the workflows and the parameter settings that must be applied during the production of a statistical output. It records all the instructions and information needed for statistical processing of a survey from collection to dissemination.

An individual workflow points to where data is stored and specifies a method or statistical process that must be applied to it. A simple workflow may consist of a datastep that derives a new variable. It specifies the data to be extracted, the code that is to be executed on the variables, and the new variables to be created. A more complicated workflow might extract a large dataset and pass it through a complex statistical tool like Banff or QualityStage. The workflow specifies all the parameters and rulesets required by the statistical tool.

A subject-matter specialist or methodologist builds a new statistical output by creating a new configuration in the configuration store of the microeconomic platform. This configuration will specify all the transformations and statistical methods that must be applied to the data to produce the output and the order in which they must be applied. The version-controlled configuration store ensures permanent traceability between raw data, the data process rules, the methods are applied, and the final outcome.

The components of a configuration can be copied easily and reconfigured for a different need. This means that processes, methods, queries, and reports that have been developed for one collection can be easily applied to another. For example, a generic editing and imputation process is available for any survey or administrative dataset by changing the instructions and parameters recorded in the configuration store. Standardisation is enforced, because new configurations are required to draw on existing standard configurations composed of processes based on standard tools.

The configuration store allows subject-matter specialists and methodologists with a basic understanding of computer logic to set up all the statistical processes needed to produce new statistical outputs with minimal IT support. The configuration for an output breaks the end-to-end process into a large number of small steps, so complex data transformation and statistical processing is achieved by combining well-tested data processes and statistical tools. An advantage of this approach is that the production system is created as the new output is being developed, assessed, and tested.

For example, the configuration store contains a generic automated edit and imputation process that can apply a wide range of edit types and all standard imputation methods to any survey or administrative dataset. To apply this process to a new dataset, the subject-matter specialist or methodologist amends the configuration to record the methods, parameters, and other information that is needed to apply the required edit and imputation process to the new data.

New methods can be developed, tested, and implemented within the existing processing system. They can be introduced to other outputs by changing the relevant configurations.

The configurations make extensive use of SAS for three reasons:

- SAS is a high performance statistical tool for data processing and statistical analysis that manages large-scale data manipulations with ease.

- A library of standardised and tested SAS-based statistical methods is already available within the organisation.
- Many subject-matter specialists are trained in SAS.

The disadvantage of SAS is that subject-matter specialists and methodologists can rewrite complex code again and again in an unstructured way, defying standardisation and efficiency. They can create new SAS datasets easily, making security and version control extremely difficult.

The microeconomic platform controls and limits the use of SAS by using the configuration store as a facility for writing and managing SAS in a structured and standardised way. The benefits of SAS described above are retained, but the pitfalls are avoided because the platform controls the way SAS is executed and securely manages the resulting data.

Although the new microeconomic platform uses SAS, it is not tied to this particular software. When using SAS, the workflow sends the code stored in the configuration store to a SAS server for processing. Other calculation engines, such as R and Excel, can also be used in the same way.

3.1.3. Scenario Store

The scenario store is a key element in an analyst-driven statistical system. Whenever an automated process to change data is applied, the entire dataset is extracted to a temporary area, called the scenario store. If the analyst selects the current approved configuration ruleset for the scenario (i.e., the version used in producing outputs) then, after confirming that results are as expected, the data can be written back into the main permanent store and from there be used in outputs. However, the subject-matter specialist can copy an approved ruleset, make changes and create an experimental scenario at any time. This experimental scenario can be used for evaluation. The approved and experimental scenarios sit side by side in the store for easy comparison using standardised views and tools.

This means that as well as changing the process rules, analysts can easily test those changes with live production data without affecting production systems in any way. In this way, existing rulesets can be continuously improved or completely replaced. The scenario area can be used to prototype entirely new outputs or to make small changes in existing rules. The rights of subject-matter specialists to make changes are managed by security. There is version control of the configuration rulesets and the scenario store. A change management process ensures that all changes are properly tested and approved before they are implemented in published statistics.

3.1.4. Reduction in Staged Processing

Reviews of existing systems to inform the new design showed that data processing was often broken down into many stages. Each stage had its own data preparation followed by manual inspection. Sometimes data is repaired manually when, if left alone, it would have been corrected by automated repair processes at a later stage. Processing analysts often had to carry out the same manual repair multiple times if data was reprocessed. If the output result was several stages away, more serious quality issues could easily be missed. Later detection meant more rework. In summary:

- Each process operated independently of the others, often in different parts of the organisation. Problems and errors arose from the transfer of data between environments.
- Inspection and manual repair was inefficient and duplicated.
- The contribution that each process made to the whole was not always clear to those responsible for designing and monitoring it. It is often not clear whether they are adding value.
- The effect of actions cannot be seen until the end of the last process is complete and an output is produced. This may be days or weeks after the first process began.
- When errors are detected in later stages, the process is very slow to repeat, or in the worst cases cannot be repeated in the time available before output deadlines.

These problems have been resolved by building larger, more holistic processes that can be run repeatedly until quality targets are met.

In this iterative approach, each time the subject-matter specialist runs the configuration that controls the statistical production process, the complete set of workflows is run over the data and the results are available for inspection via a standardised reporting suite. Over time, as more data is received in sample surveys, or rules are adjusted in administrative collections, scenarios move closer and closer to the final outcome. Approximate output results are always available. This methodology is born of necessity as the organisation moves into processing more large-scale administrative data where manual intervention and manipulation is impractical. In summary:

- Data is managed in a single very large cohesive environment.
- Processing is, as much as is possible, done in a single pass and in a single scenario set.
- Manual editing is post-process and driven by significance and by output needs.
- Processing and revision of data is easily repeatable.
- The configuration is designed and built by subject-matter specialists as a cohesive single sequential process. The configuration specifications are improved over time as they are fine-tuned by real world data experience.

Strategically, the microeconomic platform makes product quality highly visible. It also allows changes to maintain quality, for example updating an industry classification to the latest version, to be carried out quickly and cheaply. Operationally, product quality is controlled through security settings that allow only a very few people with appropriate authority to make changes to configurations used to produce outputs. Change control policies and procedures ensure proper testing and review has accompanied proposed changes.

The microeconomic platform is making statistical development a more standardised, more cost- and time-effective process. These attributes are essential for achieving the goals of the Stats 2020 programme of change.

4. Statistical Production Process

The development of the new platform has had a significant impact on the statistical production process, particularly editing and imputation. Modern designs typically include administrative data in some way, often in large quantities, and use standard concepts, methods, and tools.

A major part of the regular government audits of Stats 2020 has to do with measuring improvements. The most significant savings will be made through operating efficiencies in relation to the new statistical production platforms that have started to replace the outdated and dissimilar existing systems. The four perspectives referred to in Section 1 are the strategic goals of the organisation for the next seven years (transforming how statistics are delivered, leading the New Zealand OSS, obtaining more value from official statistics, and creating a responsive and sustainable organisation). The programme is planned to reduce labour input by just over 20% in this time frame. Each of the Stats 2020 projects has short-term performance targets, and a time frame in which expected benefits should be achieved. Early indications for the microeconomic platform are ahead of this target in both labour savings and the time taken to achieve them.

For example, a study by the methodology group found that the number of units needing manual edit intervention decreased by 66% in repeating business surveys. This was based on the use of the standard significance measure deployed in the new system. The business area has confirmed this saving in practice with operational changes. Preliminary results for another survey, using Banff to do automated repairs, combined with the impact assessment above should see realised labour savings of 20% or more.

Development of the microeconomic platform has not involved a substantial upfront investment. It has however been completed over quite a long period of time, taking four years to deliver a full suite of business processes. Delivery was deliberately planned in phases, with each phase of the programme being self-justifying in terms of costs and benefits.

The first stage focused on a large administrative data series that was not an output in its own right, but is used to replace small units in a major postal survey. This made the initial phase and goal much simpler. A target of reducing postal units in the survey by around 60% over a five-year period was achieved by using more of the administrative data, developing a more efficient data handling capacity and automated editing and imputation. At the close of the first phase, a cashflow analysis (project cost input, reduced postal costs output) showed an investment return on capital of 18%. In addition, the project had created generic data handling and edit and imputation processes that could be used for other collections.

The second phase of work added the processing of GST (or value-added tax) unit records to the platform. This phase enhanced the ability of the system to handle longitudinal data. It also increased the confidence of the organisation in the development and led to the organisation committing to further investment. These very large administrative datasets are routinely transferred electronically from collecting agencies to Statistics NZ through a standardised process. The wide variety of types of administrative data and of statistical issues encountered, such as common mistakes that occur when filling out a form, encouraged a pragmatic approach. Editing strategies aimed to resolve the most common errors in form filling, such as ignoring a request to use thousands of dollars when answering financial questions. Errors that are less common, but are material, are also resolved. For the sake of platform performance, errors that are less common and immaterial are ignored. The way statistical issues such as definitional differences and temporal or cross-sectional aggregation effects are resolved depends on their importance with respect to the statistical product being produced using techniques such as statistical modelling.

The third phase of work brought the first postal business surveys onto the platform and the very large Overseas Trade series. During this phase, the investment costs for development and implementation rose dramatically, as sample selection and survey management functionality was added to the platform. However, Statistics NZ was willing to commit to this investment, because the core elements of the platform were already in place and proven. Efficiencies were able to be achieved immediately for successive products, because the generic editing and imputation tools were already available. The platform is able to cope with a wide variety of data formats, data quality, and processing methods. In tandem with the platform development, the organisation is focussed on developing a proven set of methods and tools for evaluating, collecting, processing, analysing, and disseminating products based either entirely on administrative data or on a combination of data from multiple sources. Techniques such as statistical modelling and data integration are especially important for making extensive use of administrative records for statistical work.

The fourth phase of the project will see 19 statistical outputs moved to the platform in the next two years. For some of these outputs, new functionality will have to be developed. For example, the Business Register has been incorporated onto the platform, which has created a need for new functionality to enable handing of units with data that does not change every month, but does have complex ownership relationships that need to be maintained in real time without creating performance issues. The investment for development and implementation in this phase is greater as the platform is being scaled up from a few users to hundreds of users. This requires intensive investment in reliability and robustness of performance, with separate teams being established to support survey migration activity and provide support for the new standardised platform. One result of this is a change in focus from teams supporting individual surveys to teams supporting clusters of surveys using common processing systems and common data storage.

The project to develop a microeconomic platform was not carried out in a way that required a large investment upfront; rather, it was established and proven in stages over a long period. A clear overarching vision guided the development but clear targets were set for results in each phase. In the past, development would have been preceded by carefully focused and relevant analyses by methodology teams before a start was made on the IT requirements; these recent developments were approached in a different way that allowed for a much more pragmatic decision-making process. Standard tools were predominantly off-the-shelf rather than built in-house, so development work was more in the nature of adding tools to the platform and optimising tools for particular collections rather than starting from scratch and redesigning entire collections. All collections migrating to the new platforms are expected to adopt the available standard tools and there are very few exceptions. Teams are multidisciplinary and include methodologists, subject-matter specialists, and IT developers. IT business analysts provide a link between methodology and technology. They play a key role in ensuring that all business requirements are articulated and that the implications of requirements in terms of statistical and business architecture are understood.

Changes to collections on the new platforms are carried out incrementally in small steps rather than via a major effort in system development and implementation. Statistical redevelopment is planned and implemented in a programme of work called the Statistical Infrastructure programme, a ten-year rolling programme of statistical redevelopment to maintain outputs at the desired quality where quality is considered in all its dimensions. Examples include updating the industry codes in business surveys when the business activity classification has been updated, changes in methods, and responding to real world events such as the recent change in New Zealand's rate of GST. A key goal for this project is that all investment decisions around statistical development priorities take an organisation-wide view. Until recently, redevelopment decisions were made independently of quality and priority frameworks. A key difference between development in the past and development on the new platforms is that in the past a change would need to be applied in each affected collection separately, typically sequentially because of dependencies between collections. Now many changes to collections can be applied to all collections on the platform simultaneously.

A key to minimising production costs is the statistical toolbox, which includes standard tools for common business processes, many sourced from other national statistics offices. For example, editing and imputation uses the Banff and Canceis programs developed at Statistics Canada (Statistics Canada 2005; Bankier et al. 2001). X-12-Arima is used for seasonal adjustment, and QualityStage™ is used for data integration. In this way, the organisation promotes the use of standard tools for common processes including detecting outliers, deterministic imputation, historic imputation, donor imputation (and many other standard imputation models), seasonal adjustment, and data integration.

Tools in the toolbox are categorised according to whether they are current, emerging, legacy, or graveyard tools. Emerging tools are under evaluation for inclusion in the toolbox; current tools are those actively promoted for use within the organisation and supported by the organisation; legacy tools have been identified as needing to be replaced with more up-to-date versions; and graveyard tools are being phased out. The corporately supported tools are not always the most efficient, but they usually represent the standard methodological approach for common problems. Exceptions to the use of current tools are only permitted by agreement at the highest levels of the organisation. The tools are accessible from the platforms as a tools library that allows analysts to choose a tool and use it in the production process the analysts are creating.

Standard tools have been made available on the production platforms as the need arises – not all tools are currently available for use in the new production environments. Methodologists are still exploring some tools; for example, Statistics NZ has been evaluating the selective editing tool SELEKT developed by Statistics Sweden. However, it isn't clear yet how SELEKT could be used in a fully automated processing system. If the goal is to score records based on the significance of any change, and then review the records with the highest scores, then a simple scoring mechanism may suffice. In a number of cases, tools have been modified to be of more use. For example, a version of Banff's Proc Outlier that includes significance criteria for identifying outliers has been developed, because of the prevalence in business tax data of units that are extreme but also extremely small.

To achieve the organisation's efficiency goals, production processes for designs using administrative data are predominantly automated, as the volume of data precludes the possibility of manual data editing. Unlike data the organisation collects itself, contacting the data supplier to clarify data anomalies is not possible with administrative data. Full end-to-end automation of business processes reduces manual editing of records, improves data quality, and substantially reduces the time and staff required to produce statistics.

Adopting an analyst-driven approach has changed the underlying architecture of the microeconomic platform. In the role-based approach, handovers of data and manual intervention are minimised. Reliable large-scale processes automatically process data that is stored in a central datastore. The goal is to make it easier for analysts to build and test new configurations or make changes to existing production systems. The emphasis is on building a highly flexible, rapidly repeatable process. Rule changes and manual repairs are implemented at the end of each cycle and the whole process is repeated until the desired outcome is achieved.

Examples of cost savings and improvements in quality include:

- The three administrative data collections and two business surveys currently using the microeconomic platform are all using standard tools for each of their business processes.
- Moving to the new platform has seen either a reduction in the size of each survey or administrative data collection's processing team; or a reduction in processing time; or both.
- Analysts using the platform are able to move easily between the outputs because of the common look and feel of production systems and the use of common tools.
- Complete audit trails are available of the changes to each record and this information has allowed managers to easily review each output's business processes and make improvements in time and cost of production. For one of the first four outputs introduced to the new platform, the publication has been brought forward.
- Statistical development for those outputs already on the platform is able to be carried out without IT intervention in most cases, thereby reducing the cost of the work significantly. The New Zealand Harmonised System Classification 2007 (NZHSC2007), used in overseas merchandise trade statistics, was able to be updated to NZHS2012 by a single subject-matter specialist. Data processing and checking is a single continuous process, and all analysts access the same dataset. This has eliminated the problem of different analysts holding and using different versions of the same dataset.

The platform has a high degree of flexibility in accommodating new methodologies. The platform is able to load and process most common types of survey and administrative data. Changes to editing and imputation strategies, such as adding new edits in response to changes in the underlying data and using alternative imputation methods, are simple to test and deploy. Standard statistical tools are available for most common business processes.

5. Governance

Significant governance is key to ensuring that generic solutions are delivered and that Statistics NZ meets its organisational efficiency goals. Stats 2020 is a large and costly programme of work, depending upon the success of over 100 multi-year projects. The New Zealand government has made a substantial investment in the first four years of the programme, with additional funding of approximately NZ$60 million for operating costs over that time, primarily funding common statistical infrastructure and common platforms, methods, processes, and tools (Statistics NZ 2010). For that reason, there is more oversight of the organisation by the government and a higher level of reporting requirements than ever before.

The governance framework for the oversight and assurance of the Stats 2020 programme consists of a set of functional responsibilities that have been developed for project management, governance, and project assurance. The Stats 2020 governance framework is additional to the existing organisational governance model and is intended specifically to support the delivery of the Stats 2020 programme. Such a formal approach to monitoring the programme's progress is necessary to ensure that the organisation is well placed to report on progress to the government.

Governance bodies called portfolio committees have been established for clusters of dependent and/or related projects. The portfolio committees monitor benefit realisation, for instance that all projects in the portfolio are making their expected contributions to the strategic priorities.

Each project has a project owner and a project manager who are focused on project outcomes and project outputs respectively. Steering committees focus on the project scope, financial risks, and roadblocks; they provide operational direction for the projects and have a quality assurance role. Steering committees have the authority to make decisions on specific project components, and enforce the organisation's standards for quality and risk management.

Decisions that result in changes to a project's scope, timescales, or budget that are outside agreed project tolerances are raised to the appropriate portfolio committee for approval. The portfolio committee structure is different from the usual line management; decision making is therefore independent of the usual line management considerations and takes an OSS perspective.

Personnel are managed using a matrix-management model: line managers are responsible for the day-to-day performance and development of their staff, and project managers are responsible for the project management aspects of staff assigned to projects. All projects are prioritised annually and decisions about resource allocation take into account the priority of each project, always aiming to keep projects that have a higher priority fully resourced.

A Transformation Office provides programme assurance. The Transformation team monitors the programme from a benefits perspective and provides independent strategic oversight and advice to the Government Statistician. Key members of this team are portfolio programme managers whose role is to provide strategic oversight across portfolios, advise on progress, and identify opportunities to leverage change. In addition, a project management team has been established to provide internal project management assurance. The team provides training and support on project management and the organisation has invested significantly in project management training for all project leaders.

This governance model is supported by:

- a Standards Review Board, which oversees the corporate standards and prescribes their use.

- a statistical toolbox, which contains the current set of standard statistical tools that are to be used in systems developments.
- steering committees that provide operational direction and quality assurance to the clusters of projects.
- training for staff in how to assess the quality of administrative data, process administrative data, and develop statistical products based either entirely on administrative data or on a combination of data from different sources, and
- training for staff in how to use the standard tools and design statistical production systems (e.g., common editing and imputation strategies for survey data and administrative data, typical order of business processes, and monitoring process quality).

Performance indicators associated with each of the Stats 2020 projects show baseline and target information and an indication of the time frame in which benefits are expected to be achieved. An example of a performance target might be lower end-to-end production costs. When performance indicators measure end-to-end production costs of products, the baseline is the direct operating cost of producing each product. Direct operating costs are available from the finance system, which is detailed enough to allow direct operating costs, such as time spent preparing for a release, to be separated from costs associated with development work, since these activities are "coded" differently in the time-keeping system.

6. Lessons Learnt

The move from stand-alone production systems to a model where generic components are developed and delivered has been difficult. In the mid- to late part of the last decade, the organisation tried to implement a large-scale project known as the Business model Transformation Strategy (BmTS). It aimed to create an organisation-wide framework of standard processes and systems for statistical production. The BmTS was designed to achieve this through a programme of projects – each project delivering different components of a complete end-to-end processing system (Savage 2008). For example, one project would deliver an input data environment for physical data storage; another project would deliver an output data environment for data dissemination.

The Legolution system is one example of a processing system developed using BmTS principles (Savage 2008). It was developed in 2005 to process a suite of surveys measuring business performance (e.g., the Research and Development Survey). Legolution provided generic data editing and imputation, using Banff, a data editing and imputation system developed by Statistics Canada (Statistics Canada 2005).

Many of the principles that have been adopted for the microeconomic statistics platform had been tried out in the BmTS-based systems. For example, every output processed in the Legolution system had a documented data editing and imputation plan, which included measures of processing quality to enable monitoring of the system and allow continuous improvement. Unfortunately, the project had to be de-scoped during the initial development because the work involved in creating a generic solution applicable to more than one survey had been underestimated and, as a result, the solution was not as generic as first specified.

The BmTS project struggled to deliver some key elements within the time, scope, and budget constraints imposed by user needs. Ultimately the programme was cancelled (Savage 2008). The short time frame for delivery was at odds with the large scale of the components being developed. Also, building generic components in separate projects resulted in incompatible solutions. The generic lost out to the specific, causing affected projects to reduce their scope and diverge in order to meet the immediate business goals of each project. At the time, it seemed that the drive for standardisation was lost. However, the ideas and lessons learnt had been carried forward:

- The BmTS goals of standardisation and an integrated data environment are kept in mind but reprioritised to make business delivery the focus and generic features more staged.
- Development teams adopt agile software development techniques (e.g., Highsmith 2004). The system is developed and delivered incrementally so that analysts are able to conduct tests very early on and their feedback is used to improve the product. Prototypes are developed using real data to ensure the final product is fit for purpose.
- Subject-matter specialists and methodologists adopt a more hands-on and involved approach, becoming part of the development team and involved in all aspects of testing and delivery. The team responds to change rather than working to plan. The incremental approach allows for changing user needs while minimising costs.

Now, seven years on, the goals of the BmTS are beginning to be realised through the microeconomic statistics platform.

Providing subject-matter specialists with generic highly automated production systems is not enough to change an organisation's culture. Traditionally, Statistics NZ has had what is known as a *process* culture: there was little or no feedback and people focused on doing things right rather than on doing the right things. The organisation was a typical hierarchical government department where predictable and consistent systems and procedures were highly valued and people had delegated authorities within a highly defined structure.

In this process culture, decisions about statistical maintenance of outputs were made independently of prioritisation and quality frameworks. The decisions were driven by the need to replace aging production systems. Because they were reactive rather than planned for, IT systems developments were exceedingly expensive and, because they happened at widely spaced intervals, resulted in production silos, unique production systems for each survey run by Statistics NZ where any further changes to systems (e.g., updating classifications) could only be made by IT teams. Different production systems for each survey meant it was difficult for people to move between surveys without a significant amount of training and analysts became experts in processing data and troubleshooting IT systems rather than in data analysis.

The new platforms allow statistical maintenance to be carried out in small regular increments and use standard tools for common processes. The advantage of the change to a cluster approach (i.e., groups of surveys and administrative data collections processed on the same IT platform using the same tools for editing, imputation, confidentiality, data dissemination, and analysis) is that analysts have more time to concentrate on data analysis. When statistical maintenance is required, the new platform is highly configurable so changes are easy to implement and only minimal IT intervention is needed. For example, a new classification can be introduced to the platform by a subject-matter specialist. All they need to do is to change the metadata and business rules in the configuration store. The impact of the classification changes can be tested in the scenario area without any risk to existing statistical outputs.

It has not been easy to move from a process culture to a more constructive culture – where quality is valued over quantity, creativity is valued over conformity, cooperation is believed to lead to better results than competition, and effectiveness is judged at the system level rather than the component level (Cooke 1987). Statistics NZ is still in the early stages of achieving this transformation. The organisation has developed a strategy to help teams understand and explore what the new environment means for them. The people strategy has been broken down into two areas of work:

- a workforce programme, which focuses on building on people's skills and capabilities, and makes sure systems are in place to attract, develop, and appropriately reward people, and
- a culture and change leadership programme, which focuses on supporting people through change, developing strong leaders, and building a performance culture.

A key enabler of the culture change to date has been the adoption of an agile project management approach for IT development projects. The agile approach involves setting up multidisciplinary teams, planning the team's work in manageable pieces that take a short period of time to complete (e.g., two weeks), and then checking in with the team regularly to monitor progress and remove impediments. Progress is very fast under this approach and impediments to progress are identified and resolved quickly. While developed primarily as an iterative method of determining requirements in engineering development projects, it can be used in any team situation and has had the added benefit at Statistics NZ of blurring role boundaries, encouraging all team members to help solve any problem. Team capability develops very quickly and the team environment becomes more conducive to the creation of innovative solutions.

7. Conclusion

Statistics NZ aims to produce statistics that are fit for purpose in a cost-efficient and sustainable way. With the Stats 2020 programme of work, the organisation is developing and implementing business architecture that will support the organisation's goals. Statistics NZ is sharing the knowledge and experience gained with other national statistics offices both informally and formally, through the international collaboration network instigated by the Australian Bureau of Statistics in 2010.

The move from stand-alone systems to large-scale IT platforms is underway. The large-scale platforms will support the processing and analysis of multiple outputs and are based around standard business and information concepts encapsulated in standard methods and systems. Opportunities identified during the move include providing systems built in such a way that subject-matter specialists can easily create new or modified outputs by applying standard statistical tools and processes, or combining data from different sources, without the help of IT teams. The use of standard tools and processes ensures that production systems are automated as fully as possible. Governance structures, a people strategy, and a change leadership programme focused on moving towards a more constructive culture and supporting people through change are in place to support it.

8. References

Bankier, M., Poirier, P., and Lachance, N. (2001). Efficient Methodology Within the Canadian Census Edit and Imputation System (CANCEIS). Proceedings of the Annual Meeting of the American Statistical Association, August 2001. Available at: http://www.amstat.org/sections/srms/proceedings/y2001/Proceed/00202.pdf (accessed December 2012).

Biemer, P.P. (2010). Total Survey Error: Design, Implementation, and Evaluation. Public Opinion Quarterly, 74, 817–848.

Bycroft, C. (2011). Social and Population Statistics Architecture for New Zealand. Available at: http://www.stats.govt.nz/~/media/Statistics/surveys-and-methods/methods/research-papers/topss/Social%20and%20population%20statistics%20architecture-jul11.pdf (accessed December 2012).

Cooke, R.A. (1987). The Organizational Culture Inventory. Plymouth, MI: Human Synergistics, Inc.

Glaude, M. (2008). A Strategy for the Development of Social Statistics in the European Statistical System. Eurostat. Available at: http://unstats.un.org/unsd/statcom/statcom_08_events/special%20events/New_directions_social/Michel_Glaude_Paper.pdf (accessed December 2012).

Groves, R.M. and Lyberg, L. (2010). Total Survey Error: Past, Present and Future. Public Opinion Quarterly, 74, 849–879.

Highsmith, J.A. (2004). Agile Project Management: Creating Innovative Products. Boston, MA: Pearson Education.

Kent, J. (2011). Availability of Data from Scarcity to Profusion. Paper presented at UNECE Conference of European Statisticians, Geneva. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2011/wp. 33.e.pdf (accessed December 2012).

McKenzie, R. (2008). Statistical Architecture for a New Century. Paper Presented at International Seminar on the Use of Administrative Data in Daejeon, Korea. Available at: https://www1.oecd.org/std/41142889.pdf (accessed December 2012).

Metis Steering Group (2012). The Generic Statistical Business Process Model. Available at: http://www1.unece.org/stat/platform/download/attachments/8683538/GSBPM+Final.pdf (accessed December 2012).

Savage T. (2008). Statistics New Zealand's Move to Process-oriented Statistics Production: Progress, Lessons and the Way Forward. Paper presented at the IAOS Conference 2008: Smart Data, Innovative Uses – Reshaping Official Statistics in Shanghai, China, October 2008. Available at: http://isi.cbs.nl/iaos/conferences/Papers/CS_19_2_Savage_Gretton.doc (accessed December 2012).

State Services Commission (2011). Performance Improvement Framework: Formal Review of Statistics NZ. Available at: http://www.ssc.govt.nz/sites/all/files/pif-statisticsnz-review-oct11.pdf (accessed December 2012).

Statistics Canada (2005). Functional Description of the Banff System for Edit and Imputation V1.04. Ontario: Statistics Canada.

Statistics NZ (2010). Statistics New Zealand Strategic Plan 2010–20. Available at: http://www.stats.govt.nz/~/media/Statistics/about-us/corporate-publications/strategic-plan/snz-strategic-plan-2010-20-new.pdf (accessed December 2012).

Statistics NZ (2012). Integrated Data Infrastructure and Prototype. Available at: http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/integrated-data-infrastructure-and-prototype.aspx (accessed December 2012).

Thomsen, E. (2002). OLAP Solutions. Building Multidimensional Information Systems. New York: Wiley.

Received December 2011
Revised July 2012
Accepted November 2012

Journal of Official Statistics, Vol. 29, No. 1, 2013, pp. 99–124

Addressing Disclosure Concerns and Analysis Demands in a Real-Time Online Analytic System

Tom Krenzke[1], *Jane F. Gentleman*[2], *Jianzhu Li*[1], *and Chris Moriarity*[2]

This article focuses on methods for enhancing access to survey data produced by government agencies. In particular, the National Center for Health Statistics (NCHS) is developing methods that could be used in an interactive, integrated, real-time online analytic system (OAS) to facilitate analysis by the public of both restricted and public use survey data. Data from NCHS' National Health Interview Survey (NHIS) are being used to investigate, develop, and evaluate such methods. We assume the existence of public use microdata files, as is the case for the NHIS, so disclosure avoidance methods for such an OAS must account for that critical constraint. Of special interest is the analysis of state-level data because health care is largely administered at the state level in the U.S., and state identifiers are not on the NHIS public use files. This article describes our investigations of various possible choices of methods for statistical disclosure control and the challenges of providing such protection in a real-time OAS that uses restricted data. Full details about the specific disclosure control methods used by a working OAS could never be publicly released for confidentiality reasons. NCHS is still evaluating whether to implement an OAS that uses NHIS restricted data, and this article provides a snapshot of a research and developmental project in progress.

Key words: Confidentiality; data utility; disclosure risk; statistical disclosure control.

1. Introduction

To meet data user needs, a battery of systems for facilitating access to data need to be provided. The battery of systems that a government survey organization offers to facilitate data access must be integrated, not just to maximize usefulness and convenience to users,

[1] Westat, 1600 Research Boulevard, Rockville, MD 20850, U.S.A. Emails: tomkrenzke@westat.com; janeli@westat.com

[2] National Center for Health Statistics, 3311 Toledo Road, Hyattsville, MD 20782, U.S.A. Emails: jgentleman@cdc.gov; cmoriarity@cdc.gov

Acknowledgments: The authors of this article are employees of NCHS and of Westat, NCHS' contractor for investigating, developing, and evaluating disclosure avoidance methods for possible use in a future online analytic system for NCHS survey data. The UCLA Center for Health Policy Research is also serving as a subcontractor. The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the National Center for Health Statistics, Centers for Disease Control and Prevention. The authors gratefully acknowledge the many individuals who are contributing to the project. We thank NCHS staff, including Brenda Baker, our original contract Project Officer, for her attention to this project; Susan Jack, who provided valuable assistance on the extensive preparations of variables in the underlying microdata test files; and Jacqueline Lucas, who assisted in providing the necessary microdata test files. At Westat, we are grateful to Michele Harmon, our Project Director, for her oversight on this project, and to Craig Ray for his leadership and experience relating to system development. At UCLA, special thanks to AJ Scheitler and Alberta Soranzo for sharing their experience and providing their input on the interface design. We are also indebted to the Integrated Health Interview Series staff at the University of Minnesota for providing advice in variable selection and assistance with classifications for variables used in NHIS test files. Lastly, the authors are grateful for the special methods review conducted by Natalie Shlomo.

but also for disclosure avoidance. Maintaining confidentiality (which is mandated by law and which is critical to continuing survey respondent cooperation) is becoming increasingly complex as data user sophistication and access to technology increase. This article describes how the National Center for Health Statistics (NCHS), the official health statistics agency of the United States federal government, is developing methods that would be required to offer to the public an interactive, integrated, real-time online analytic system (OAS) to facilitate analysis of both restricted and public use survey data. Data from NCHS' National Health Interview Survey (NHIS) are being used to investigate, develop, and evaluate such methods. Detailed information about the NHIS is available at: http://www.cdc.gov/nchs/nhis.htm.

An OAS must be integrated with existing means of data access and must also within itself deal with the tradeoffs between providing access to data and observing the agency's own rules and laws for protecting confidentiality. The NHIS website releases annual public use NHIS microdata files as well as documentation and numerous analytic reports (NHIS 2010, 2011). Like all of the surveys conducted by the NCHS, the NHIS has a mandate to protect the confidentiality of survey participants. Personally identifiable information (PII), such as persons' names, is not released. Certain "restricted variables," including U.S. state identifiers, lower-level geographical identifiers, and sensitive variables, are not available in public use files (PUFs) because the combination of low-level geographic data and demographic data on an individual microdata record could lead to a breach of confidentiality. In general, this means that detailed versions of indirect identifying variables are not released to the public. By "indirect identifying variables," we refer to variables that are factual in nature, such as age, gender, and geographical area, and that can reveal the identity of an individual when used in combination.

The NCHS alleviates limitations on public access to its restricted data by providing access to NCHS restricted data through its Research Data Center (RDC) (NCHS RDC 2011; Meyer, Robinson and Madans 2012; http://www.cdc.gov/rdc/).

The NHIS was designed mainly to produce national estimates. However, the NHIS collects data in all 50 states and the District of Columbia, and the NHIS sample is designed so that NHIS data are representative of the target population (the civilian noninstitutionalized population) at the state level as well as at the national level. As discussed in Gentleman (2012), analysis of state-level data is appropriate and desirable for studying the U.S. health care system because the U.S. health care system is largely administered at the state level. Nevertheless, appreciable numbers of state-level NHIS sample sizes are normally too small for analysts to produce stable state-level estimates from a single year of data. Sometimes, satisfactorily precise state-level estimates can be produced by combining two or more adjacent years of data. However, for confidentiality reasons, NCHS does not include any state identifiers on the NHIS PUFs. Thus, when NHIS state sample sizes are insufficient for a particular state-level analysis, an NHIS data user from outside NCHS who wishes to perform that analysis must currently utilize the NCHS RDC.

In developing a new policy to provide access to restricted microdata, an agency must identify and consider what types of data access already exist (such as through an RDC) so that new uncontrolled or inadvertent risks to confidentiality are not created. In particular, the disclosure avoidance methods needed for an OAS are highly dependent on whether or not PUFs are also produced, because users of an OAS can compare its results with PUFs and analyses thereof.

The NHIS has been in the field since 1957. Since 1997, the Family Component of the NHIS has collected data about persons of all ages in each family within each interviewed household, and additional data have been collected about one randomly selected child (the "sample child") and from one randomly selected adult (the "sample adult") in each family. The existence of a PUF (and any other publicly available linkable file) makes the creation of an OAS for the NHIS particularly challenging, as discussed in this article.

The OAS under consideration for use with the NHIS would use NHIS data starting with the 1997 data year. It would integrate two real-time OAS subsystems in a manner that would be transparent to the user: 1) "Subsystem P," which uses only public use data; and 2) "Subsystem R," which uses its own underlying microdata files containing selected restricted data (e.g., state) as well as public use data. All developmental work assumes this system design. In addition, OAS analyses would be performed in real time with no pre-prepared results.

NCHS is still evaluating whether to implement an OAS that uses restricted NHIS data, and this article provides a snapshot of where the OAS research and development project is at the time of writing. Therefore, it is important to note that all aspects of the system discussed in this article are under evaluation, including, as examples, a utility to allow the user to combine categories of variables, the parameters for thresholds on unweighted counts to determine whether or not to deny results from being shown to the user, and the Statistical Disclosure Control (SDC) treatments.

The OAS approaches discussed here are generic for agencies/organizations that conduct household sample surveys, provide PUFs, and have confidentiality constraints. We do not attempt to explore disclosure risks related to administrative records data or to data with highly skewed distributions, such as data from establishment-based surveys.

In Section 2, we discuss OAS features being considered, including general data management for data to be accessed in the integrated system, analysis methods, and transmission and storage of the results of analyses. Section 3 presents a discussion of challenges and trade-offs that impact design decisions. Section 4 describes the main disclosure threats to such a system and investigations of how to protect against such risks. Section 5 presents various statistical disclosure control (SDC) treatments under consideration. An initial evaluation of some SDC approaches is discussed in Section 6 in terms of their impact on disclosure risk and analytic utility. Lastly, concluding remarks are given in Section 7.

2. General OAS Features

The general system architecture for an OAS such as the one under consideration for use with the NHIS would include the distinct system components that are described below and depicted in Figure 1.

User interface. For greatest efficiency and cost effectiveness, it is necessary to include the interface design as an integral part of the system development. For the OAS under development, the user interface is implemented through a web-based, statistical reporting tool, for which the challenge is to arrive at a single interface that provides a

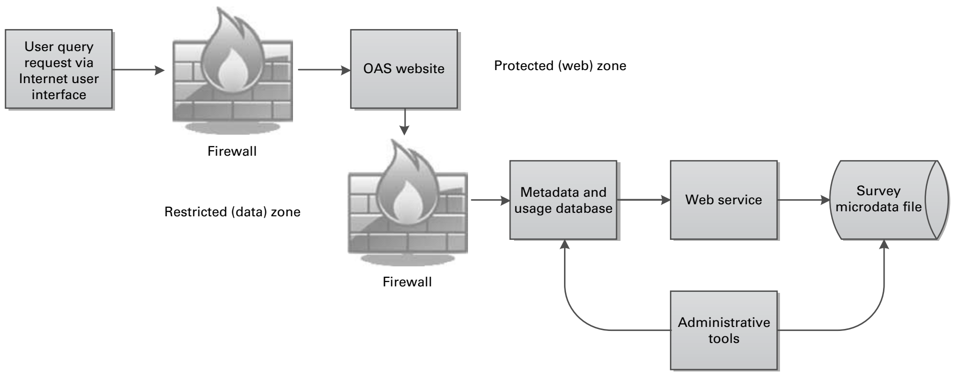

Fig. 1. General System Architecture from Query to Data Access

user-friendly environment that appeals to a wide audience. The interface would accommodate sophisticated queries as well as guide novice users effectively through the system to generate results that meet their needs.

Metadata database. A metadata database is needed to provide the capability to easily define datasets and to group variables into logical "folders," to enable selected options for statistics and missing values, and to define the contexts in which each variable can be used. A metadata database helps to determine which full-sample weight is needed to request estimates from different sub-questionnaires, to provide information about rounding rules for each variable, and to indicate the data years for which each variable is available. The metadata database is impacted by disclosure risk concerns. For example, it will contain information to help integrate the two subsystems. If a restricted variable (e.g., state) is specified in the query, then Subsystem R, its SDC treatments, and the restricted data would be used. The database would contain information on the subsetting population filter variables for Subsystem R. If a restricted variable is not being used, then Subsystem P, without restriction, would be used. System usage should be tracked within the database, including information about queries to inform system managers about what variables are most frequently requested. If the OAS database were, for example, implemented using a SQL Server, the website would communicate with the database via a metadata and usage class library.

Administrative tools. The system needs to include a suite of parameter-driven administrative tools that allow for rapid configuration of projects using the same interface and analytic engine. A parameter-driven system provides the flexibility to put in values that define certain rules. For example, rounding rules can be turned off, or the amount of rounding can be changed. Under evaluation is an administrative infrastructure to maintain user profiles and project-specific configurations, and to migrate metadata from a staging area to the production environment.

Survey microdata file. The integration of data sources would allow queries to be requested using variables from multiple sources. It may be beneficial to combine data into one large file with multiple sets of weights and sets of survey variables, and include several survey years. The survey microdata files are impacted by disclosure risk since the restricted data files may have been treated by random perturbation, while the corresponding public use files may not have been perturbed for the same variables.

Types of analysis and analytic engine. The types of analyses under consideration for the OAS include descriptive statistics (proportions, means, medians and other quantiles, etc.), multi-way cross-tabulations, ratios, standardized rates, standard errors, confidence intervals, and graphs. The graphing utility will only display estimates that have passed disclosure-related threshold rules set up for the tables (e.g., as discussed in the introduction in Section 5). More advanced statistical options could include choice of confidence interval width and type and provide the capability to do significance tests to compare pairs of cells in a cross-tabulation. Any changes to the types of analyses, such as a new approach to estimating confidence intervals or adding new modeling capabilities, might require enhancements to the analytic engine. With interface panels devoted to each components of the query (e.g., selecting geography, variables for tables, statistical options), a natural progression would be to add a new panel to the interface to devote to regression modeling capabilities. In addition, the methodologies related to reducing disclosure risk are relatively fluid and continue to develop with new approaches. For example, for regression analysis, Reiter (2003) provides an approach to perturbing residuals prior to displaying a diagnostic plot. Chipperfield and Yu (2011) provide a discussion of disclosure rules applied to regression analysis in an OAS. Once a system is launched, if an SDC treatment is enhanced or changed, it could impact on the consistency of results from prior runs. Therefore, a balance between reducing disclosure risk further and the impact on data utility would need to be considered.

A benefit is that since the analytic engine is a portable utility (i.e., un-plug and plug-in capabilities), it would allow for any methodological (statistical or disclosure-related) changes and additions to be developed in-house and uploaded without disruption to users. The analytic engine is optimized for speed of computation, and takes into account the complex sample design when estimating sampling error. A replication method to estimate variances from sample surveys with complex sample designs is being evaluated. Replicates are random subsamples of the full sample, and replicate weights are created for each replicate to account for the sample cases that were dropped from the full sample to form the subsample.

The existence of disclosure risk makes an impact on how the components of the OAS architecture are constructed. We discuss further the disclosure threats with respect to challenges faced during development in the next section, and we discuss the specific disclosure risks, various limitations and SDC treatments later in the article.

3. Beginning the Work: Challenges and Trade-offs Relating to the System Architecture

We describe here some challenges and trade-offs encountered and addressed during the early phases of development of the OAS that may give others a head start in the development of their own systems. The challenges for the architectural structure are classified into three categories: 1) balancing risk, utility, and user friendliness; 2) system issues; and 3) maximizing statistical methodology. We start by describing some decisions that had to be made in balancing risk, utility, and user friendliness.

Decision about providing ways to combine multiple years of data. The utility of Subsystem R can be increased through limited efforts of constructing consistent recodes (new variables created from existing variables), values, and value labels across data years.

Decision about what age groupings to permit. Data analysts would ideally be free to choose whatever age groupings suit their specific subject matter, but the desired age groupings, especially for health data, often are not common (e.g., five-year) age groupings. An intruding user, if allowed to choose any age groupings, could deliberately select age groups that differ by, for example, one year, which would make it possible for a differencing/linking attack to occur. The topic of differencing tables is summarized in Duncan et al. (2011), and an example in the context of official statistics can be found in ONS (2006). Striving for flexibility in defining age ranges to meet the needs of the data user led to the initial development of a dynamic utility for combining (but not disaggregating) variable categories within the OAS.

Decision about the extent to which users should be protected from denials. Various threshold rules were decided upon to deny tables of high risk. Threshold rules are discussed further in Section 5. In an ideal OAS using NHIS data, analysts would be free to request, for example, a state-level cross-tabulation involving any state. But requests for analyses of states with small populations would very likely have to be denied by the system because of unacceptably high risks to confidentiality.

Decision about the number of variables to allow in a given query. Initial investigations show that limiting the number of variables used for subsetting the data so as to define the subpopulation or creating universe of interest helps to reduce disclosure risk. Therefore, decisions were made as to the maximum number of variables to allow (i.e., about a dozen variables) and which ones were analytically important to keep for this purpose.

Decision about how much to tell users about the confidentiality protection methods used. Analysts of perturbed data need to be warned that their analyses will have reduced accuracy and increased inconsistencies due to the application of confidentiality protection methods. A summary of approaches to perturb microdata is given in Duncan et al. (2011).

The following decisions address challenges relating to the subsystems.

Decision about integrating Subsystems R and P. It was decided to integrate Subsystem R and Subsystem P because many users would not be able to or want to specify to which subsystem they want to submit their queries ahead of time. This led to the development of a smart system to determine which underlying microdata file to use (PUF or restricted), depending upon the variables in a specific query. The challenge arises in the handling of changes to the query as the user provides a developing series of specifications for the desired analysis.

Decision about the structure of the underlying files. We were faced with several options relating to file structures underlying the system. The decisions took into account the size of the file, processing of integrated public and restricted data (perturbed), and the various levels of the NHIS data (i.e., file types). Because a query that uses restricted data must use perturbed data, it was decided to make a set of the files containing just the public use data and a set containing perturbed restricted use data. This simplified the system processing at the expense of higher storage costs.

The following decisions address challenges while attempting to maximize the statistical methodologies available.

Decision about dynamic combining of categories. Without the ability to combine categories, the user is constrained by the categorized versions of the variables in the underlying dataset, with no ability to modify the variables. However, the combining of variable categories increases disclosure risk, which must be attended to in developing the system. The associated increase of disclosure risk due to dynamic combining of categories is touched upon in Section 6.1.1. In addition, the decision to offer dynamic collapsing of categories has several implications for creating estimates for a variety of statistics.

Decision pitting confidentiality against consistency. In an OAS using NHIS data, analysts using Subsystem R to obtain state-level estimates will likely want to compare those state estimates to national estimates. Using Subsystem R's underlying microdata file to compute both state-level and national estimates would provide consistency. For smaller domains, intense exploration by an intruder of the differences between estimates produced from Subsystem R data files and the same estimates produced from PUFs could result in a threat to confidentiality. To avoid such intrusion, the current approach is for Subsystem R to use PUFs to calculate any requested estimate that does not involve restricted variables (e.g., a national estimate). Under that plan, some inconsistencies caused by differences between Subsystem R's underlying data file and the PUFs will need to be tolerated.

Decision about rounding rules. Rounding reduces the amount of detailed and precise information. As an example, the NCES Statistical Standards Section 5.3 (NCES 2003) addresses rounding rules that should be applied to numbers and percentages in text and summary tables.

Decision about dynamic degrees of freedom. Degrees of freedom are used in calculating confidence intervals and conducting significance tests. They are an indication of the stability of the estimated sampling variance. In the case of an OAS using NHIS data with a focus on providing state-level estimates, and because some states have a small number of primary sampling units, the associated degrees of freedom could vary widely. The dynamic approximation of the degrees of freedom in the OAS depends on the subgroup, number of active replicates, the type of replication approach (which may require the number of unique strata in the subgroup), and the type of statistic.

4. Challenges Relating to Disclosure Risk in an OAS

A major task is controlling risks involved within the data that can be exposed in a real-time OAS. The main disclosure risk elements in an online table generator are 1) cross-tabulations with small unweighted cell sizes, especially in the margins of the tables, 2) table linking, 3) weights that vary and variables with a large number of categories, and 4) table differencing. We expand on each of these risks.

There are different ways to define small cell sizes. By "cell size," we refer to the underlying unweighted number of respondents from which the weighted cell frequency was calculated. Sometimes the "Rule of Three" is used based on the concept of k-anonymity (Sweeney 2002), which specifies that at least three individuals must be present (a count of at least three). The Rule of Three is based on the premise that a singleton (a table cell that is based on one sample case) can possibly be identified by another individual, while a cell count of two is susceptible to disclosure if a sampled individual finds himself in the cell and deduces the characteristics about the other individual in the cell. Singletons and doubletons (table cells that are based on one and two sample cases, respectively) are typically high risk. A singleton is sometimes referred to as a "sample unique." Sample uniques are not necessarily unique in the population. Ways to find risky data values by estimating the probability of being a population unique are discussed throughout the literature and summarized in Hundepool et al. (2012).

Table linking is an attacking technique used to string together information, called the "matching key," from several tables in order to form a microdata record on a singleton. By "matching key," we refer to a combination of variables from the restricted file that, when used together, can be matched to a publicly available external file in order to merge restricted variables (e.g., detail on geographic location) to the public file. This type of table linking in the context of producing tables (albeit static, as opposed to real-time in this article) when a PUF exists is described in the context of the Census Transportation Planning Products (CTPP) on five-year American Community Survey (ACS) data (Krenzke et al. 2011b) and the existence of the ACS public-use microdata sample (PUMS). For the OAS, when a singleton exists in the marginal of variable S_0 (e.g., a questionnaire or registry variable such as gender) in cross-tabulations with other variables, and S_0 appears in several explicit table queries, a microdata record can be created for the singleton. Suppose a series of two-way tables (where a "*" means "crossed with") using other questionnaire or registry variables is requested, such as S_0*S_1, S_0*S_2,. . . ,S_0*S_3, and suppose that a category in S_0 is comprised of a singleton. The tables can be linked together to define a microdata record for the single sample unit consisting of values of S_0, S_1, S_2, S_3,. . . ,S_k, for k variables crossed pairwise with variable S_0. Suppose the series of tables is generated for a subgroup R=1 defined by a restricted variable (e.g., State=Maryland). This may allow restricted information, such as detailed geographic information, to be assigned to a public-use record on an external file.

Fienberg (2009) summarizes various disclosure risks attributable to weights. Because sampling weights almost always vary, tables with many cells that are generated from a small universe usually would show unique weighted totals for selected cells that would show up for other tables requested for the same universe, leaving the system susceptible to table linking. By "universe" we refer to a key domain of interest to make inference about, which could be a group or subgroup of persons specified by the user, and based on one or more variables. Suppose S_0 has two levels (such as gender), and there are three records with $S_0=1$. The following series of tables (Table 1) shows the sum of weights for variable S_0 being crossed pairwise with S_1 (six levels), then with S_2 (eight levels), and then with S_3 (four levels), for the restricted subgroup R=1. Other variables can be crossed pairwise in the same manner with S_0 but are not shown here. Moreover, only the row for $S_0=1$ is shown in the illustration.

Suppose the sampling fraction is such that a weight of 1,000 is one of the smallest weights on the file. As one example from the series of tables in Table 1, an intruder can see a very small number 1,000 for $S_0=1$ and $S_1=1$, then 1,000 for $S_0=1$ and $S_2=1$, and lastly 1,000 for $S_0=1$ and $S_3=2$. Therefore, the intruder can link the tables together using the common variable S_0 and determine that $S_1=1$, $S_2=1$, and $S_3=2$. Since 1,000 continues to appear in the tables, the intruder would grow in confidence that the cell is a singleton. Even when two sample records fall into the same cell, as shown in the third part of Table 1 where $S_0=1$ and $S_3=1$, one knows from the prior tables that the two contributing weights are 3,000 and 3,500, since those are the only two that can sum to 6,500. With this intruder approach, the tables can be linked using their sampling weight to arrive at pseudo-microdata records displayed in Table 2, which would serve as a "matching key" to attempt to attach the restricted subgroup to the public file.

Table 1. Series of tables illustrating table linking due to weights for subgroup R = 1

	S_1					
S_0	1	2	3	4	5	6
1	1,000	3,000	0	0	0	3,500
2						

	S_2							
S_0	1	2	3	4	5	6	7	8
1	1,000	3,000	0	0	0	0	0	3,500
2								

	S_3			
S_0	1	2	3	4
1	6,500	1,000	0	0
2				

Differencing tables from multiple requests poses another challenging threat. By "differencing," we refer to an intruder attack where the difference is taken between common cell estimates between two explicit tables displayed in the online tool in order to create an implicit table (perhaps written down on paper). Suppose there are two tables that use the same table variables but differ in the universe specification. Suppose both queries successfully satisfy query threshold rules (such as the Rule of Three, which requires three or more sample cases). These two queries may lead to an implicit table (the difference between the two tables) that would have failed the query threshold rule if explicitly specified. This is also sometimes referred to as "slivering," since the multiple requests may create "slivers" of universes.

Likewise, differencing tables from multiple requests on the same universe can lead to the derivation of tables on subgroup slivers when "not applicable" populations occur for items in different tables. Given the above risk elements, we formulate the intruder scenario as it would relate to Subsystem R. We define the following groups of variables:

R = Set of restricted variables that would reside on Subsystem R only.
D = Set of demographic variables and other indirect identifiers that reside on the public use file (PUF) and within Subsystem R. The source of the variables could come from a registry or the questionnaire. Examples are age, race, sex, and marital status.
S = Set of questionnaire variables that reside on both the PUF and within Subsystem R, which contain subject matter information, such as body mass index (BMI), duration

Table 2. Illustrated microdata formed by table linking due to weights

Pseudo-microdata record	S_0	S_1	S_2	S_3
1	1	1	1	2
2	1	2	2	1
3	1	6	8	1

variables, limitation variables, and so on. These are variables that an intruder would likely not know about the individual.

O = Set of other variables that are not available to Subsystem R, but are on the PUF. These variables would include any variables suppressed from Subsystem R. Some variables may be included in the PUF only. For example, if the PUF offers national-level analysis and R offers state-level analysis, the questions that were answered by a very small subset of participants (e.g., due to skip patterns) can only be analyzed at the national level. Since these data are so sparse, collapsing or recoding would not work for these variables. Another example is that Subsystem R may include the collapsed version of a variable, but the more detailed versions are in the PUF only.

L = Set of variables that are not on Subsystem R, nor on the PUF, but are on other files that can be linked to Subsystem R. An example is in reference to in-house files and those files that are accessible through an RDC.

Subsystem R variables can be represented by R||D||S, where "||" indicates the concatenation or union of the groups of variables. The PUF variables can be represented by D||S||O. All publicly available data that are available through linking IDs on the public files may contain at most the following D||S||O||L, where D, S, O, and L are disjoint. Some variables in L are only available in the RDC's restricted files through Designated Agent agreements. Here we want to be clear that we are not expanding the space of alternative data sources to include nonpublic sources. The situation is illustrated in Figure 2.

A data intruder may try to conduct multiple queries on a real-time OAS using restricted data, using table differencing, in order to link implicit "slivers" (from differencing) to obtain the "matching key" (D||S) for a limited number of records contained in the sliver. The matching key can be used to try to attach R to the PUF, which would be a violation of the PUF data use restrictions and may lead to the identity of an individual, or at least increase the chance for inferential disclosure. The set of variables identified by this can be represented by R||D||S||O. Certainly, then, it is simple to expand the set by linking other linkable files, which can be represented by R||D||S||O||L. In summary, it follows that an objective of the SDC treatment is to address table differencing (TD) to prevent a microdata record from being formed in a real-time OAS using restricted data through linking implicit tables (LIT), and then conducting a record linking (RL) operation to put restricted data on the PUF and other linkable files. Thus, we refer to this as the TD-LIT-RL

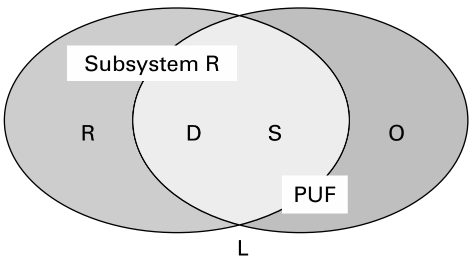

Fig. 2. Schematic representation of survey variables and associated files

(or differencing/linking) sequence attack, which comprises the highest chance of success in any intruder action. In such a case, in the absence of SDC procedures, it would be highly likely to have a one-to-one match with the existing NHIS public use files.

5. SDC Treatments

For an OAS, an additional risk is the availability of a PUF, and therefore in general, an objective of the SDC treatments is to protect all variables offered in the OAS in some manner, since any subset of the variables can be used in a record linking operation. The approach for Subsystem R of the NHIS OAS currently under review consists of one-time SDC treatments on the underlying microdata, and modification of analytic results in real time within the OAS' Subsystem R.

The purpose of treating the underlying microdata is to add uncertainty to the identification of an individual in the sample. The following steps would be taken to achieve that goal:

- Initial risk analysis. The goal of the initial risk analysis was to identify high risk data values to inform the data coarsening and random perturbation processes.
- Data coarsening. By "data coarsening" we mean global recoding, including top-coding (trimming high outlier values to a cutoff) and variable suppression. A discussion of data coarsening can be found in the *mu-Argus* 4.2 manual (Statistics Netherlands 2008), and summarized in Hundepool et al. (2012).
- Random perturbation. By "random perturbation" we imply an SDC treatment used to modify data values using a controlled approach with a random mechanism. Some perturbation approaches are discussed in FCSM (2005). Two examples of random perturbation approaches are data swapping and the constrained hotdeck. Data swapping procedures have been discussed throughout the literature (e.g., Dalenius and Reiss 1982; Fienberg and McIntyre 2005). Data swapping is used at the U.S. Census Bureau, where in practice pairs of households are swapped across different geographic regions, as discussed in Ramanayake and Zayatz (2010). Controlled random swapping is described in Kaufman et al. (2005) as it relates to National Center for Education Statistics data. The constrained hotdeck is a relatively new approach among data replacement methodologies that was developed through research for the National Academy of Sciences (Krenzke et al. 2011a).
- Evaluation of the impact on data utility. Reviewing weighted frequencies and reviewing associations between variables before and after random perturbation are typically done. As an example for multivariate associations, Woo et al. (2009) proposed using propensity scores as a global utility measure for microdata.

Within Subsystem R, it would be of extreme importance that proper confidentiality screening be conducted and that the system architecture be constructed to handle it. That is the most critical and the most challenging requirement. In addition to the research being done at NCHS (Gentleman 2011), other agencies and organizations have attempted or are attempting to address the disclosure risk elements of a dynamic OAS that utilizes restricted data, such as the UCLA Center for Health Policy Research's AskCHIS (UCLA 2012), with limited suppression and rounding, the U.S. Census

Bureau (Lucero et al. 2011), with subsampling and threshold rules, the Australian Bureau of Statistics (ABS) (Tam 2011), which applies a post-tabular perturbation approach, Statistics Canada (Simard 2011), which applies random rounding, and the U.S. Substance Abuse and Mental Health Services Administration, which also has disclosure limitations embedded in it. By "post-tabular perturbation," we refer to either a random rounding mechanism or a look-up table that contains the noise to add to the original estimate. In some respects, the SDC approaches under consideration for Subsystem R are influenced by the approach being implemented in the Census Bureau's developing Microdata Analysis System (MAS), discussed in Lucero et al. (2011) and Freiman et al. (2011). Like the scenarios being considered for Subsystem R, the MAS plans contain rules to restrict universe and table query specifications, apply threshold rules, and include a subsampling approach referred to as the "Drop Q Rule."

5.1. Goals of SDC Treatments in Subsystem R

A general goal when applying SDC treatments to developing an OAS in the presence of a PUF is to limit distortion in the results while protecting against the TD-LIT-RL attack. With this in mind, we define the following properties when developing an OAS:

(1) Cell consistency – Across multiple users, if the same set of records contribute to a table cell, the same results are attained.
(2) Query consistency – Across multiple users using the same query path (e.g., same specification for universe definition and requested table), the same results are attained.
(3) Additivity – The sum of results from multiple tables is equal to the results directly arrived at for the aggregated table.

It is interesting to note that through our risk evaluations, we have found that attaining cell consistency may not provide adequate protection against an extensive TD-LIT-RL attack under extreme conditions when a PUF exists. For instance, in the extreme scenario of table differencing (TD) for explicit tables that differ by one case, attaining cell consistency results allows for the potential identification of the true attribute since all cells but one have zero sum of weights in the implicit tables from differencing. Therefore, when linking implicit tables (LIT), a set of true attributes are revealed and can be used in record linkage (RL) to a PUF, for example.

Attaining query consistency helps to provide credibility in the results. With such a property however, it must be understood that two users who strive for the same estimate through two different paths may get different results.

Attaining additivity, while achieving consistency, is difficult in a dynamic query system. The Australian Bureau of Statistics (Fraser and Wooten 2005) achieves consistency in the first stage through a post-tabular adjustment, and then applies an additivity algorithm in the second stage using a form of iterative proportional fitting (IPF), which was introduced by Deming and Stephan (1940). The consistency property is lost during the second stage since the IPF is applied to the specific table at hand.

The task is, however, challenging to a lesser degree for surveys considering an OAS without a PUF. When a PUF does not exist, the focus is typically on reducing the risk for a limited number of indirectly identifying variables (the ones that provide factual information about an individual). As mentioned previously, when a PUF exists, all variables must be protected because it becomes a computer matching operation if a microdata record can be formed through the TD-LIT-RL attack. Table 3 provides a summary of SDC treatments that were considered for a real-time OAS using restricted data.

There are four main SDC treatments being considered:

(1) Query restrictions.
(2) Threshold rules.
(3) Dynamic subsampling.
(4) Rounding.

The implementation of the approaches in Subsystem R (query restrictions, threshold rules, subsampling, and rounding) took much collaboration among systems staff, computer programmers, statisticians, and program managers. It was necessary that each approach be fully understood, and many times trade-offs were discussed in terms of costs, amount of effort, processing time, and complexity.

5.2. Query Restrictions

Once within the system, the user would specify the items of interest: main topic, geography, universe, and table request. Based on the query, the system would be able to determine if the query should run through Subsystem P or Subsystem R. Within Subsystem R, query restrictions would be employed to reduce disclosure risk. For example, only one restricted variable could be used in a query. As another example, after an evaluation of risk, it was determined that one of the largest risk-reducers was to only make a limited number of “filter” variables available to define a subpopulation. The systems architecture issues relating to query restrictions included how to identify the variables that were allowed to serve as filter variables, and to be able to manage the flow between screen panels as the system tries to determine which subsystem to invoke.

5.3. Threshold Rules

The system would check whether the query would pass threshold rules, such as those given below.

Example Rule 1. Universe Threshold Rule. The purpose of the Universe Threshold Rule is to prevent small cells from occurring in the subsequent table request after the initial request. This rule denies a query result if the specified universe has less than X records, where X is a prespecified and confidential integer.

Example Rule 2. Table Marginal Count Threshold Rule. The purpose of the Table Marginal Count Threshold Rule is to prevent table linking from explicitly specified tables. This rule denies a query request if the specified m-way table has less than Y records in any $m-1$ marginal, where Y is a prespecified and confidential integer. The cell means or quantiles need to be based on at least Y records in order to help reduce the risk of disclosing specific values of the continuous variable. The systems architecture issues relating to threshold rules were straightforward to address. If the query fails the threshold

Table 3. Examples of possible SDC treatments in a real-time OAS using restricted data

SDC Treatment	Risk-reducing benefit	Utility impact	Performance
Limiting scope of queries (e.g., not allowing more than three variables in a cross-tabulation)	Disallows attempts by intruder to drill down into the details of the microdata file	Reduces the scope of analysis	Negligible impact
Threshold rules	Denies tables with small sample sizes to an intruder attempting a table linking attack	Denies tables with small sample sizes to the user	Negligible impact
Subsampling, Post-tabular perturbation, Post-tabular random rounding	Provides a level of protection against a differencing/linking attack, depending upon the approach	Has potential to retain consistency; however, the additivity property is lost. Impact on variance estimates would be readily retained in subsampling	Initial tests show that the impact of subsampling on performance is minimal. Impact of post-tabular approaches minimal
Post-subsampling calibration to grand total for the universe	Provides additional level of protection against a differencing/linking attack	Reduces the difference for universe totals. May have some negative impact for table cells, depending on how it is applied	Initial tests show that impact on performance due to an extra step depending on the table generator is minimal
Rounding output	Provides some protection against a differencing/linking attack	Reduces the detail in the estimates displayed to the user	Negligible impact

rules, a message would be displayed on the screen to provide some general guidance to the user.

5.4. Dynamic Subsampling

Dynamic subsampling developed for Subsystem R is influenced by the U.S. Census Bureau's Drop Q Rule approach and the post-tabular perturbation approach of the Australian Bureau of Statistics (ABS) described in Fraser and Wooten (2005).

The following approach to subsampling within an OAS achieves query consistency and limited additivity, while protecting against a TD-LIT-RL attack. Suppose a two-way table is specified. Suppose U is the sum of random numbers (RND) for records in the defined universe. Suppose M1 is the sum of RND for records in the row marginal, M2 is the sum of RND for records in the column marginal, C is the sum of RND for records in the table cell of interest. Let SEED be a function of U, M1, M2, and C. The derived seed is a function of the query request (defined universe and table variables) such that the same seed would be generated for the selection of the same subsample within a table cell.

The dynamic subsampling occurs cell by cell; that is, the resulting subsample for the query is stratified by the table cells. The number of cases to deselect is a function of the cell's sample size. The seeding approach reduces the risk of an intruder making an accurate inference after taking the difference between two tables generated on slightly different universes.

Prior to computing the results, the sampling weights (full sample and replicates) are adjusted within table cells to account for the cases not selected in the subsample. This ensures that the weighted totals are unbiased and that the variances are computed properly. After the within-cell reweighting, another overall reweighting step is used to calibrate to the grand total. Regarding systems architecture, the system needs to first detect that the query should invoke Subsystem R. The analytic engine is programmed to conduct the subsampling and reweighting, and return the results to the interface. Initial test runs indicate that the implementation of dynamic subsampling does not result in a measurable increase in the run time.

5.5. Rounding Rules

Rounding rules can be used for the resulting estimates to reduce the chance for linking tables with weighted totals in cells with small sample sizes. The following are examples under consideration:

- Weighted percentages – nearest 10th (0.1).
- Standard errors and confidence intervals of weighted percentages – nearest 100th (0.01).
- Weighted counts – nearest thousand (1,000).
- Standard errors and confidence intervals of weighted counts – nearest hundred (100).

If the statistics of interest are means or percentiles, the rounding rules in Table 4 are being considered.

The systems component most impacted by the rounding is the metadata database. The database provides rounding rule information to the NHIS OAS for each variable in the

Table 4. Rounding rules for means or percentiles

Analysis variable	Rounding rule for means, medians	Rounding rule for standard errors and confidence intervals
If 0 − 1 variables	0.001	0.0001
Else if 25th percentile ≤ 9	0.01	0.001
Else if 25th percentile ≤ 99	0.1	0.01
Else if 25th percentile ≤ 999	1	0.1
Else if 25th percentile ≤ 9,999	10	1
. . .	. . .	. . .

underlying microdata file. The analytic engine receives the rounding rule information, which depends on the query that is specified, including the variables involved and the statistics requested. The rounding rules (RR) for the corresponding precision measures such as standard errors and confidence intervals can be derived accordingly (dividing RR by 10). If the statistic of interest is the weighted total of an analysis variable, we consider using RR(100) (meaning, round to the nearest multiple of 100) as the rounding rule for the total. If the statistic of interest is the ratio of weighted totals of two analysis variables, we consider using RR_{num}/RR_{denom} as the rounding rule for the ratio, where RR_{num} is the rounding rule associated with the variable in the numerator, and RR_{denom} is the rounding rule associated with the variable in the denominator.

5.6. Limitations of the SDC Approach

We note the following limitations of the SDC approach. First, due to the perturbation approach in data preparation, there is some increase to the mean square error term. Second, the property of additivity will not be preserved due to missing data, the subsampling (or post-tabular perturbation) approach, and rounding. That is, two tables added together may not equal the aggregate if estimated directly. To address this issue, the tables would need to be estimated directly, and not as the sum of individual tables.

6. Analysis of the SDC Impact on Risk and Utility

The impact of the set of SDC treatments that were discussed in the previous section was evaluated in terms of the competing objectives of reducing disclosure risk while retaining data utility. Here we provide a step-by-step approach to the evaluation such that the procedures are emphasized, while the outcomes of the evaluations are still under review and need to be kept confidential.

As discussed previously, the biggest challenge to a real-time OAS using restricted data is to protect against a TD-LIT-RL attack. This occurs when implicit tables with small cells (from table differencing) can be linked together to form a string of identifying characteristics (“matching key”). The matching key can be matched to external databases, such as a corresponding PUF. Recognizing that combinations of just a few variables can lead to a singleton, it was necessary that the initial risk evaluation of the microdata take into account the ability to set system parameters, such as only allowing up to three-way tables, or only allowing a certain level of geography. The main objectives were to determine the query restrictions, threshold rules, the parameters relating to dynamic subsampling, and the rounding rules.

6.1. Impact on Disclosure Risk

Here we take the risks in turn with a discussion of table differencing (TD) and linking implicit tables (LIT), and then record linkage (RL).

6.1.1. TD and LIT Risk

To reduce the risk due to table differencing and linking implicit tables, the following Subsystem R treatments are being evaluated: query restrictions (e.g., the number of variables allowed to define the universe and tables), threshold rules, dynamic subsampling, and rounding rules.

Query restrictions and threshold rules. To evaluate the impact from query restrictions and threshold rules, restricted NHIS test microdata files were used for six combinations of data years (2009 and 2007–2009) and file types (Person (Family Core), Sample Adult, and Sample Child). Restricted NHIS data were accessed at the NCHS Research Data Center in Hyattsville, Maryland. While the sample size varies across the years, the 2009 Person File had about 88,000 records, which was about four times as much as the Sample Adult File and about eight times as much as the Sample Child File. The six combinations of years and file types were processed first through an initial risk analysis.

Exhaustive tabulations were processed for a given set of variables, which formed m-way tables. Given lower (MINDIM) and upper (MAXDIM) bounds for the table dimension size m, n tables were formed by

$$n = \sum_{m=MINDIM}^{MAXDIM} \binom{p}{m} = \sum_{m=MINDIM}^{MAXDIM} \frac{p(p-1)\cdots(p-m+1)}{m(m-1)\cdots 1} \tag{1}$$

where p is the number of variables used in the table generation. For example, suppose there are $p = 20$ total variables and all two-way, three-way, and four-way tables are created, then 6,175 tables would be generated. If the number of cases in a cell was less than the threshold, the cell was flagged as a violation cell, and the variables/categories used to define the cell were identified as "contributing to cell violations." Violation counts were the number of times a record was in a violation cell. The percentage of cell violations for a variable/category was computed as follows:

$$V = 100[V_c/k_c] \tag{2}$$

where V_c is the number of violation cells involving this variable/category c, and k_c is the total number of cells formed by this variable/category c. The above formulation can be adapted to define universes in lieu of defining tables. The records were classified into a risk measure with five categories (values 0 through 4), from low to high, based on the frequency that a record was flagged for violations. Category 0 was assigned to the records with the lowest risk, and Category 4 was assigned to records with the highest risk as determined by the violation counts.

The analysis compared different Universe Threshold Rules and Table Marginal Count Threshold Rules. Examples of variables used in the initial risk analysis were state, Core-Based Statistical Area size and status, urban/rural status, income groups, poverty status, employment status, earnings groups, living quarters, family type, education attainment, house ownership status, gender, race/ethnicity, age, country of birth, citizenship, health insurance information, height and weight of sample children, smoking status, and diabetes status, among others. We note that the number of variables allowed to define the universe, or table, had a large impact on the number of violations. The violation percentage was far greater when three variables were used, and far less with one or two variables. Surprisingly, combining years impacted the number of violations to a lesser extent. In general, threshold rules provided much protection – without rules, tables would show small sample sizes or singletons, and table linking could occur to form microdata records.

It is typically assumed that an intruder does not know if a given person is in the sample, in which case the probability of disclosure for a sample unique is a function of the sample selection rate and the indirectly identifiable data. As an alternative to the above risk measure, many have considered the expected number of correct matches to a population file for a sample unique to be written as:

$$GlobalRisk = \sum_{SU}(1/F_k|f_k = 1) \tag{3}$$

where SU is the set of sample uniques, f_k is the sample frequency in cell k, and F_k is the population frequency in cell k. The aforementioned *mu*-Argus 4.2 manual provides a discussion of how disclosure risk is measured as a function of the sampling fraction, when an intruder knows that the unweighted cell count is one (sample unique). Investigation into the use of models to provide more stable estimates of risk has been conducted by researchers (e.g., Skinner and Shlomo 2008). A more extensive risk measure that takes into account the measurement error (e.g., data swapping) is given in Shlomo and Skinner (2010).

When determining the level of disclosure risk a file may pose, other factors need to be considered in conjunction with any single risk measure, such as the degree of potential harm to individuals if data were disclosed and the presence of risk-reducing factors, such as the level of geographic detail; availability of external data; age of the data; existence of imputation, perturbed data, recodes; and variable suppression.

Dynamic Subsampling. The benefit of dynamic subsampling is displayed in the following example. Suppose an intruder would like to make a table using a sparse category 1 of a three-category variable (A) as shown in Table 5. The intruder would be denied a table due to the threshold rule. However, if the system allows combined categories, then the categories could be combined in different trials: combining categories A=1 with A=2 and then A=1 with A=3 in separate queries. These explicit tables that pass the threshold rules could be differenced to arrive at a table that involves category 1. The first two tables show explicit two-way tables $A' \times X$ and $A'' \times X$, followed by the implicit table from the difference of the two explicit tables. For each cell, the noise added to the cell estimate due to subsampling is provided by a_{ij} and a'_{ij} where i is the row and j is the column. The noise from the subsampling is in unpredictable amounts, and therefore the intruder cannot determine with certainty the value of X in the implicit difference table even though there is a single case with a weight of 4,000 where A=1 and X=1. This is because the a_{ij} and a'_{ij} do not necessarily cancel out since the sum of RND is different in the row marginal, and the amount of the difference between a_{ij} and a'_{ij} does not have a defined limit.

Evaluations were conducted on the data to provide indications of the levels of a_{ij} and a'_{ij} under different subsampling scenarios. To understand the level of risk reduction through

Table 5. Illustration of dynamic combining of categories with subsampling

	X=1	X=2	X=3	X=4
A′: A=1,2	$36{,}000+a_{11}$	$42{,}000+a_{12}$	$16{,}000+a_{13}$	$5{,}000+a_{14}$
A′: A=3	$21{,}000+a_{21}$	$36{,}000+a_{22}$	$42{,}000+a_{23}$	$8{,}000+a_{24}$
A″: A=2	$32{,}000+a'_{11}$	$42{,}000+a'_{12}$	$16{,}000+a'_{13}$	$5{,}000+a'_{14}$
A″: A=1,3	$25{,}000+a'_{21}$	$36{,}000+a'_{22}$	$42{,}000+a'_{23}$	$8{,}000+a'_{24}$
A=1	$4{,}000+a_{11}-a'_{11}$	$a_{12}-a'_{12}$	$a_{13}-a'_{13}$	$a_{14}-a'_{14}$
A=1	$-4{,}000+a_{21}-a'_{21}$	$a_{22}-a'_{22}$	$a_{23}-a'_{23}$	$a_{24}-a'_{24}$

subsampling, we found it best to incorporate all SDC treatments used by the system (i.e., subsampling, rounding, thresholds), and then evaluate the reduction in disclosure risk through extreme examples where two universes differ by only one case. This allowed us to make clear observations about the impact on risk reduction. Through such effort, we found that when selecting a subsample in an OAS using simple random sampling there is a danger of creating a sliver. Through further evaluation, we also determined that limiting the variables to form universes greatly limits the ability to create implicit slivers with a sample size of 1.

We also note that we evaluated a post-tabular random rounding approach, and if relied upon without subsampling or other post-tabular perturbation, we are skeptical about its protection in the extreme case of a single case with A = 1 and X = 1. In this extreme case, suppose ε is the resulting amount added or subtracted to result in a rounded value of a weighted cell frequency. Suppose cells are rounded to the nearest 1,000. When table differencing (TD) explicit tables that differ by one case, the results allow a possible identification of the true attribute since all cells but one would have a difference of 1,000, 0, or −1,000. If the direction of the rounding is only triggered by the sum of the RND values for the cell, in order to retain cell consistency, then all cells in the implicit difference table, with the exception of one cell, would have a difference equal to 0. Therefore, when linking implicit tables (LIT), a set of true attributes can be revealed, which then can be used in record linkage (RL). The amount of protection is related to, and limited to, the size of ε relative to the magnitude of the estimate.

Table 6. Illustration of dynamic combining of categories with subsampling and rounding applied

	X = 1	X = 2	X = 3	X = 4
A′: A=1,2	$36{,}000+a_{11}+b_{11}$	$42{,}000+a_{12}+b_{12}$	$16{,}000+a_{13}+b_{13}$	$5{,}000+a_{14}+b_{14}$
A′: A=3	$21{,}000+a_{21}+b_{21}$	$36{,}000+a_{22}+b_{22}$	$42{,}000+a_{23}+b_{23}$	$8{,}000+a_{24}+b_{24}$
A″: A=2	$32{,}000+a'_{11}+b'_{11}$	$42{,}000+a'_{12}+b'_{12}$	$16{,}000+a'_{13}+b'_{13}$	$5{,}000+a'_{14}+b'_{14}$
A″: A=1,3	$25{,}000+a'_{21}+b'_{21}$	$36{,}000+a'_{22}+b'_{22}$	$42{,}000+a'_{23}+b'_{23}$	$8{,}000+a'_{24}+b'_{24}$
A=1	$4{,}000+a_{11}-a'_{11}$ $+b_{11}-b'_{11}$	$a_{12}-a'_{12}$ $+b_{12}-b'_{12}$	$a_{13}-a'_{13}$ $+b_{13}-b'_{13}$	$a_{14}-a'_{14}$ $+b_{14}-b'_{14}$
A=1	$-4{,}000+a_{21}-a'_{21}$ $+b_{21}-b'_{21}$	$a_{22}-a'_{22}$ $+b_{22}-b'_{22}$	$a_{23}-a'_{23}$ $+b_{23}-b'_{23}$	$a_{24}-a'_{24}$ $+b_{24}-b'_{24}$

Rounding rules. Deterministic rounding, as illustrated in Table 6, has a risk-reducing effect on the TD-LIT-RL attack. As discussed above, for each cell, the amount added to the cell estimate due to subsampling is provided by a_{ij} and a'_{ij}, and we introduce b_{ij} and b'_{ij} to represent the amount added or subtracted due to the rounding mechanism. The amount of rounding depends on the original cell estimate and the amount of noise added through subsampling. Therefore, as illustrated in Table 6, the b_{ij} and b'_{ij} do not necessarily cancel out.

6.1.2. Record Linkage (RL) Risk

Attempts were made to provide indications of the level of disclosure risk that remain for a possible sample unique record that may leak through the threshold rules, possibly through linking tables due to unusual variation in the weights across queried tables, for example.

Conditional on threshold rules being applied, and under the assumption that a series of table queries were linked to form a microdata record, the risk due to the matchability to the PUF was assessed. For the group of records in the highest risk stratum j determined through state-level tabulations using the Subsystem R data (R||D||S), we use the key D||S to attach R to the PUF. We expect E_{Mj} records to match if there was a one-to-one match, where E_{Mj} is the number of Subsystem R records in risk stratum j. When the match occurs, let A_{Mj} equal the actual number of records on the PUF (national level) that are matched, including all one-to-many matches. Without perturbation or subsampling, then the general risk measure for the highest risk records in risk stratum j was computed as $R_J = E_{Mj}/A_{Mj}$. This assessment for a group of risky records defined by the initial risk analysis provided a general indication of the risk level.

6.2. Evaluating Subsystem R SDC Treatment Impact on Disclosure Utility

To analyze the impact of subsampling on data utility, an extensive analysis was conducted on a wide variety of subsampling approaches. The evaluation included an exhaustive tabulation of state-level two-way tables using selected variables, and the relative difference between the full sample and the subsample estimates was computed. The results of each run used the same rounding rules. Both universe and marginal threshold rules were applied to suppress tables with too few cases in the universe or in any margins. The average subsampling rates within table cells dropped gradually down to a specified rate as the unweighted cell sizes increased. The results can be reviewed in risk-utility charts, such as the one illustrated from one subsampling scheme as presented in Figure 3. The chart summarizes the results from more than 1,500 tables that contained over 20,000 cells. In the graph, the 10th, 50th, and 90th percentiles of the relative differences in weighted counts are displayed at three levels: grand total, one-way tables, and two-way tables. The size strata of the subgroups, one-way tables, and two-way tables are labeled, with 1 indicating the smallest size (higher risk) and 6 indicating larger size (lower risk). Since a calibration adjustment was done to ensure that the subsample sums to the same weighted total as the full sample, the relative differences at the grand total level are zero for universes of all sizes. For two-way table cells, the ratio of full sample to subsample estimates has a median at 1 regardless of the cell sizes. The 10th

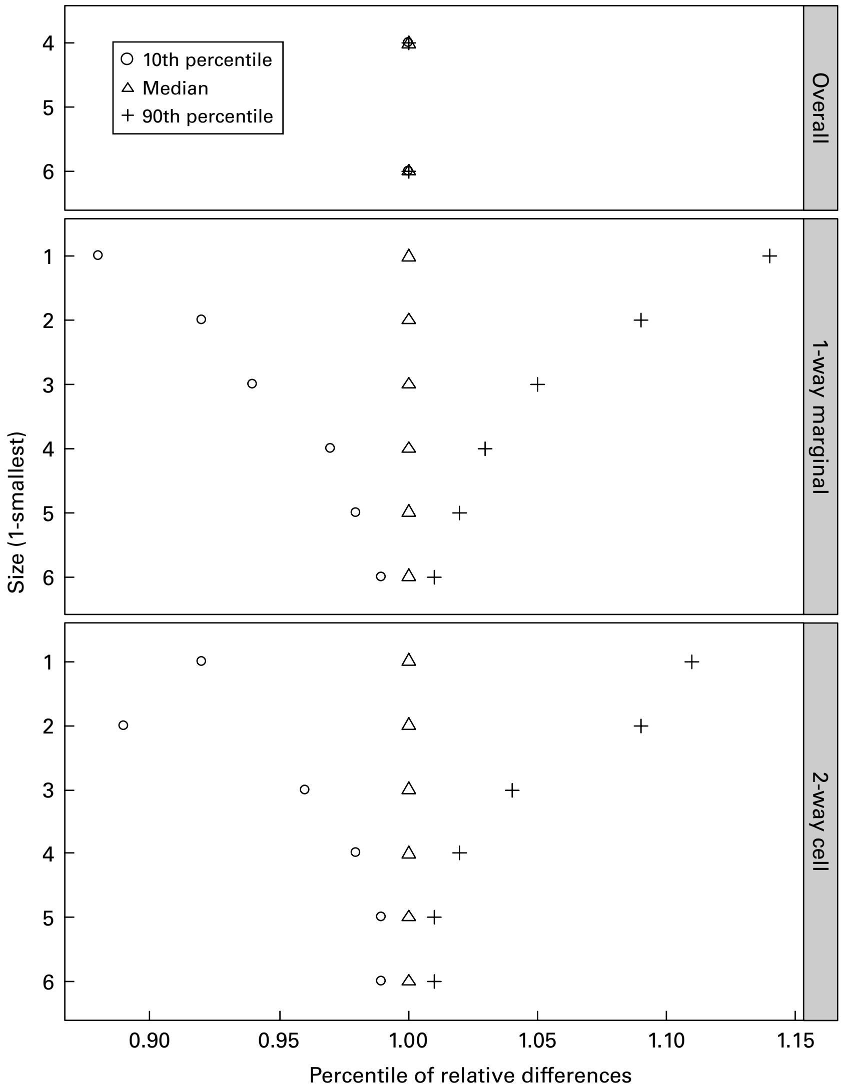

Fig. 3. Percentiles of ratios of full sample to subsample estimates by dimension and size

and 90th percentiles of the ratios show that subsample estimates for smaller cells deviate more from the full sample estimates due to higher subsampling rates.

In addition, the impact of subsampling on standard errors can be simply computed as the square root of the reciprocal of the relative difference in sample size. Table 7 provides information on the increase to the standard errors, based on the number of records deselected (q), relative to the number in the original sample (n).

6.3. Other Remarks Related to the Evaluation

If the risk measures indicate that risk levels are not acceptable, as determined by the agency, then the perturbation rates, subsample rates, and other treatments, such as

Table 7. Increase in standard errors when deselecting q records from a sample of size n

Sample size	Number of deselected records		
n	$q = 1$	$q = 10$	$q = 20$
50	1.02%	11.80%	29.10%
100	0.50%	5.41%	11.80%
200	0.25%	2.60%	5.41%
2,000	0.03%	0.25%	0.50%
20,000	0.00%	0.03%	0.05%

rounding rules, will need to be adjusted. Another approach is to be more restrictive; that is, adjust the threshold parameters to deny more tables, reduce the number of filter variables allowed, and reduce the number of variables allowed when defining tables. If the utility measures indicate that deviations are outside of tolerance bounds, then the parameters for the SDC treatments need to be reviewed with the objective of reducing the amount of the deviation. For example, the amount of data perturbation or subsampling could be reduced. If correlations indicate that attenuation is occurring, then the perturbation approach would need to be reviewed and adjusted by involving key variables in the process. As one can see, trade-offs between risk and utility cannot be avoided.

Once a system is up and running, the performance of the system needs to be evaluated. Performance of the system should be measured in terms of the time it takes from the submission of the query to the display of the results, assuming good Internet connectivity. The implementation of SDC treatment methodology, with the addition of dynamic subsampling and/or calibration, may or may not add to the server processing time due to the computations required for any one query. The extent to which the additional processing is actually noticeable to the user should be measured. If the incremental time for each query web request is excessively long, then improvements to performance may be attained by reducing the file size, reworking the SDC code, or upgrading the server platform.

7. Concluding Remarks

Two general objectives of conducting a survey are for the data to be useful and for the data to be used. As technology advances, however, the challenges of preserving confidentiality while at the same time maintaining data access increase. This article has described how we, like other researchers around the world, are striving to adapt systems architecture to facilitate methods for making restricted survey data more accessible for analysis while preserving confidentiality. The development of an OAS is a very large undertaking, but such a system has the very worthwhile potential to simplify and increase accessibility and use of the data.

Several components of systems architecture are impacted by disclosure-limitation methodology. The user interface is impacted because the interface is dynamic to the query being submitted. For example, if a state is selected for a query, then Subsystem R is invoked and fewer and sometimes different variables are available for the analysis. Since there are differences between what is allowed in a query for restricted data and for PUF data, much thought was needed on how to change the presentation of the screen contents as the user goes back and forth freely in the OAS. The system must be able to adapt to changes that the user makes to the query (e.g., deselecting a state), which can change the operating subsystem (R or P). The metadata database will contain information to help integrate the two subsystems. The administrative tools are impacted by disclosure risk, such that they control the rules relating to the protection necessary for different types of data, such as public and restricted use. The structure of the underlying microdata files is also impacted by disclosure risk if restricted data are randomly perturbed.

This article will aid others developing similar real-time online systems using restricted data, for which demands from data users are increasing. This paper provides future system developers with descriptions of tasks and issues to consider when defining capabilities and estimating costs and level of effort. For those considering an OAS as an option for disseminating survey results, including results based on restricted data, we describe the following experiences and lessons learned.

- If the home organization (e.g., a statistical agency) is not large, it will probably be necessary for it to enlist the services of a contractor to help develop an OAS.
- In implementing a new OAS, there will likely be numerous regulatory obstacles to overcome: It is difficult to implement a new system of a type that has never before been used in one's agency. Paradoxical situations may arise where conflicting demands are being made or where needs never before encountered must be met, and not everyone will fully comprehend the objectives of building an OAS. Issues may not be straightforward to address, such as where and by whom the system will be housed, hosted, and maintained.
- In implementing a new OAS, there will be numerous administrative obstacles to overcome: for example, the system developers are likely more expert in science than in contracting, despite having taken required training in dealing with contracts. Contracts will likely have to be amended and extended because of the many unknowns in the development of an OAS system.
- If more restricted data and/or new capabilities are to be added to an OAS on a periodic basis, this will require expertise that may not be found and retainable within the home organization; the knowledge and experience needed to apply specialized disclosure avoidance methods to updated data files and new system capabilities may not be available in a home organization that is not large.
- An OAS is not likely to be quickly and easily adaptable for use on a distinctly different set of data from the one for which it was originally developed.
- It is helpful for the system developers at the home organization to work very closely with the contractor to ensure that both groups have the same objectives and to obtain input from both groups in the areas of their specific expertise.

NCHS believes that online analytic systems that use restricted data can decrease the need for research data centers by making some analyses of restricted data available in a much more timely and convenient way, but such online analytic systems might also increase the use of research data centers by generating interest of the OAS users for more analyses of the type that must be conducted in a research data center. Because the future implementation of an OAS that uses NHIS data is still being considered by NCHS, it is not possible at this time to determine the effect of such a system on the overall accessibility and use of NHIS data. However, we are aware that similar systems, such as AskCHIS and the Integrated Health Interview Series' OAS (IHIS 2012), have many users, and we also hear from NHIS data users, in particular from policy makers, that they would appreciate and benefit from quicker and easier access to analyses of NHIS restricted variables as well as to analyses of NHIS public use variables.

8. References

Chipperfield, J. and Yu, F. (2011). Protecting Confidentiality in a Remote Analysis Server for Tabulation and Analysis of Data. Joint UNECE/Eurostat work session on statistical data confidentiality (Tarragona, Spain, 26–28 October 2011). United Nations Economic Commission for Europe (UNECE). Conference of European Statisticians, European Commission Statistical Office of the European Union (Eurostat).

Dalenius, T. and Reiss, S. (1982). Data-swapping: A Technique for Disclosure Control. Journal of Statistical Planning and Inference, 6, 73–85.

Deming, W.E. and Stephan, F.F. (1940). On a Least Square Adjustment of a Sampled Frequency Table When the Expected Marginal Totals Are Known. Annals of Mathematical Statistics, 11, 427–444.

Duncan, G.T., Elliot, M., and Salazar-González, J. (2011). Statistical Confidentiality: Principles and Practice. New York: Springer.

FCSM (2005). Report on Statistical Disclosure Methodology. Statistical Policy Working Paper 22 of the Federal Committee on Statistical Methodology, 2nd version. Revised by Confidentiality and Data Access Committee 2005, Statistical and Science Policy, Office of Information and Regulatory Affairs, Office of Management and Budget. Available at: http://www.fcsm.gov/working-papers/SPWP22_rev.pdf (accessed October 21, 2012).

Fienberg, S.E. (2009). The Relevance or Irrelevance of Weights for Confidentiality and Statistical Analyses. Journal of Privacy and Confidentiality: Vol. 1: Iss. 2, Article 4. Available at: http://repository.cmu.edu/jpc/vol1/iss2/4 (accessed February 1, 2013).

Fienberg, S.E. and McIntyre, J. (2005). Data Swapping: Variations on a Theme by Dalenius and Reiss. Journal of Official Statistics, 21, 309–323.

Fraser, B. and Wooten, J. (2005). A Proposed Method for Confidentialising Tabular Output to Protect Against Differencing. Paper presented at the November 2005 UNECE Work Session on statistical data confidentiality.

Freiman, M., Lucero, J., Singh, L., You, J., DePersio, M., and Zayatz, L. (2011). The Microdata Analysis System at the U.S. Census Bureau. Proceedings of the Joint Statistical Meetings, American Statistical Association Section on Survey Research Methods.

Gentleman, J.F. (2011). A Real-Time Online System for Analyzing Restricted Data from the U.S. National Center for Health Statistics' National Health Interview Survey. Proceedings of the 58th World Statistics Congress of the International Statistical Institute. Available at: http://isi2011.congressplanner.eu/pdfs/650208.pdf (accessed February 1, 2013).

Gentleman, J.F. (2012). Using the National Health Interview Survey to Monitor the Early Effects of the Affordable Care Act. Proc. 2012 Joint Statistical Meetings, San Diego.

Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E.S., Spicer, K., and de Wolf, P.-P. (2012). Data Access Issues. In Statistical Disclosure Control. Chichester, UK: John Wiley & Sons. doi: 10.1002/9781118348239.ch6.

IHIS (2012). Minnesota Population Center and State Health Access Data Assistance Center, Integrated Health Interview Series: Version 5.0. Minneapolis: University of Minnesota, 2012. Available at: http://www.ihis.us (accessed November 23, 2012).

Kaufman, S., Seastrom, M., and Roey, S. (2005). Do disclosure controls to protect confidentiality degrade the quality of the data? Proceedings of the Joint Statistical Meetings, American Statistical Association Section on Government Statistics. Alexandria, VA: American Statistical Association.

Krenzke, T., Li, J., Judkins, D., and Larsen, M. (2011a). Evaluating a Constrained Hotdeck to Perturb American Community Survey Data for the Transportation Planning Products. Proceedings of the Joint Statistical Meetings, American Statistical Association Section on Survey Research Methods.

Krenzke, T., Li, J., Freedman, M., Judkins, D., Hubble, D., Roisman, R., and Larsen, M. (2011b). Producing Transportation Data Products from the American Community Survey that Comply with Disclosure Rules. Washington, DC: National Cooperative Highway Research Program, Transportation Research Board, National Academy of Sciences.

Lucero, J., Zayatz, L., Singh, L., You, J., DePersio, M., and Freiman, M. (2011). The Current Stage of the Microdata Analysis System at the US Census Bureau. Proceedings of the 58th World Statistics Congress of the International Statistical Institute. Available at: isi2011.congressplanner.eu/pdfs/650103.pdf (accessed February 1, 2013).

Meyer, P., Robinson, E.S., and Madans, J. (2012). Protecting Confidentiality in a Data Enclave. Statistical Journal of the IAOS, 28, 25–30.

NCHS RDC (2011). National Center for Health Statistics Research Data Center website link http://www.cdc.gov/rdc/ (accessed February 1, 2013). National Center for Health Statistics, Centers for Disease Control and Prevention. U.S. Department of Health and Human Services.

NCES (2003). NCES Statistical Standards. NCES 2003-601. Washington, DC: 2003. Available at: http://nces.ed.gov/statprog/2002/std5_3.asp (accessed February 1, 2013).

NHIS (2010). National Health Interview Survey Description document. Available at: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2010/srvydesc.pdf (accessed February 1, 2013). Division of Health Interview Statistics, National Center for Health Statistics. Centers for Disease Control and Prevention. U.S. Department of Health and Human Services.

NHIS (2011). National Health Interview Survey website link, http://www.cdc.gov/nchs/nhis.htm (accessed February 1, 2013). National Center for Health Statistics, Centers for Disease Control and Prevention. U.S. Department of Health and Human Services.

ONS (2006). Review of the Dissemination of Health Statistics: Confidentiality Guidance. Health and Care. London, UK: Office for National Statistics.

Ramanayake and Zayatz (2010). Balancing Disclosure Risk with Data Quality. Research Report Series (Statistics #2010-04). Statistical Research Division, U.S. Census Bureau, Washington, D.C.

Reiter, J. (2003). Model Diagnostics for Remote-Access Regression Servers. Statistics and Computing, 13, 371–380.

Shlomo, N. and Skinner, C. (2010). Assessing the Protection Provided by Misclassification-Based Disclosure Limitation Methods for Survey Microdata. Annals of Applied Statistics, 4, 1291–1310.

Simard, M. (2011). Real Time Remote Access at Statistics Canada: Development, Challenges, and Issues. Proceedings of the 58th World Statistics Congress of the International Statistical Institute. Available at: isi2011.congressplanner.eu/pdfs/650096.pdf (accessed February 1, 2013).

Skinner, C.J. and Shlomo, N. (2008). Assessing Identification Risk in Survey Microdata Using Log-linear Models. Journal of American Statistical Association, 103, 989–1001.

Statistics Netherlands (2008). μ-ARGUS version 4.2 User's Manual. Available at: http://neon.vb.cbs.n1/casc/Software/MuManual4.2.pdf (accessed February 1, 2013).

Sweeney, L. (2002). Achieving k-anonymity Privacy Protection Using Generalization and Suppression. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10, 571–588.

Tam, S. (2011). On-line Access of Micro-Data at the Australian Bureau of Statistics – Challenges and Future Directions. Proceedings of the 58th World Statistics Congress of the International Statistical Institute. Available at: isi2011.congressplanner.eu/pdfs/650030.pdf (accessed February 1, 2013).

UCLA (2012). AskCHIS website: http://ask.chis.ucla.edu/main/default.asp (Accessed February 1, 2013). University of California, Los Angeles Center for Health Policy Research.

Woo, M., Reiter, J., Oganian, A., and Karr, A. (2009). Global Measures of Data Utility for Microdata Masked for Disclosure Limitation. The Journal of Privacy and Confidentiality, 1, 111–124.

Received January 2012
Revised March 2012
Accepted November 2012

Journal of Official Statistics, Vol. 29, No. 1, 2013, pp. 125–145

A Potential Framework for Integration of Architecture and Methodology to Improve Statistical Production Systems

John L. Eltinge[1], Paul P. Biemer[2], and Anders Holmberg[3]

This article outlines a framework for formal description, justification and evaluation in development of architectures for large-scale statistical production systems. Following an introduction of the main components of the framework, we consider four related issues: (1) Use of some simple schematic models for survey quality, cost, risk, and stakeholder utility to outline several groups of questions that may inform decisions on system design and architecture. (2) Integration of system architecture with models for total survey quality (TSQ) and adaptive total design (ATD). (3) Possible use of concepts from the Generic Statistical Business Process Model (GSBPM) and the Generic Statistical Information Model (GSIM). (4) The role of governance processes in the practical implementation of these ideas.

Key words: Adaptive Total Design (ATD); Evolutionary Operation (EVOP); Generic Statistical Business Process Model (GSBPM); paradata; survey quality, cost and risk; stakeholder utility; total survey error.

1. Introduction: Systems Development, Implementation and Maintenance in the Context of Total Survey Design

1.1. Definition and Evaluation of Systems for Statistics Production and Their Architectures

Over the past two decades, large-scale production of official statistics has become increasingly complex. Efforts to manage the resulting business processes and statistical production systems have led to development of a literature on "architecture" for these systems. Ideally, an architecture provides a reference point (used between and within statistical organisations) for efficient coordination of common business processes, information flows, production components, and development decisions. Architecture is considered especially important for reduction of development and maintenance costs and risks, for improvement of data quality, and for management of transitions from old to new production systems. See, for example, Dunnet (2007), Penneck (2009), Braaksma (2009), Pink et al. (2010), Sundgren (2010), Finselbach (2011), and UNECE (2011). In addition, a system perspective has helped some statistical organizations to improve standardization in

[1] Bureau of Labor Statistics, Postal Square Building, 2 Massachusetts Avenue, NE Washington DC, U.S.A. Email: Eltinge.John@bls.gov
[2] RTI, Research Triangle Park, NC 27709-2194, U.S.A. Email: ppb@rti.org
[3] Statistics Sweden, Box 24300, SE-701 89 Örebro, Sweden. Email: anders.holmberg@scb.se
Acknowledgments: This article benefited from many discussions with Mike Hidiroglou and several other participants in the 2011 International Total Survey Error Workshop and the Statistics Canada Symposium 2011. In addition, Boris Lorenc provided invaluable contributions to the content and exposition of this article.

order to reduce system complexity, and to identify redundant production systems. See, for example, Sundgren (2007), Field (2009), Renssen and Van Delden (2008), Renssen et al. (2009), Penneck (2009), Jug (2009), Gloersen and Saeboe (2009) and Studman (2010). Work with system architecture has also strengthened international collaborations through the use of common frameworks. See Vaccari (2009), Todorov (2011), ABS (2012), Hamilton and Tam (2012a, 2012b), and UNECE (2012a). European Commission (2009), Statistics Canada (2012), UNECE (2011), Sundgren (2011) and Munoz-Lopez (2012) provide additional background on statistical architecture; and Sowa and Zachman (1992), CIOC (1999, 2001), Rozanski and Woods (2005), OMB (2009) and The Open Group (2009) provide background on architectural frameworks and architecture of systems in general. In addition, the articles in this special issue of the Journal of Official Statistics discuss some important aspects of architecture for statistical production systems.

The abovementioned references use somewhat varying definitions of the terms "statistical production system" and "architecture." For the current article, "statistical production system" will include systems used for sample surveys (for which the statistical organization exercises a high degree of control) and for statistical work with administrative record systems and registers (for which the statistical organization may have a lesser degree of control). In addition, the term "architecture of a statistical production system" (or simply "architecture" for short) will mean an overall description of the system structure (input, output and logic), its components (subsystems), their interrelationships, and the performance characteristics of the components, the integrated system, and system outputs. Inclusion of the performance characteristics will be crucial if the architecture is to be useful to assist in improving a production system. For system output these would typically be an overall account of system capability and ability to deliver what is required by the business owner (ideally the same as the clients/users want). Section 1.2 below considers performance assessments that contribute to the architecture. These assessments simultaneously depend on methodology factors as well as factors of system design and development.

Careful descriptions of systems architecture generally require balanced discussion of features related to management, methodology and information technology. It is important that they contain views from all of these fields and that the descriptions are complementary and compatible. For some purposes, descriptions at a very high conceptual level can be useful, but practical evaluation and implementation generally require a substantial amount of technical detail. These details often involve interactions among heterogeneous subsystems, as well as evaluation of quality, cost, and risk factors over the full life cycle of the system (Camstra and Renssen 2011; Munoz-Lopez 2012). Important system features include efficiency, reliability, durability, maintainability, vulnerability, flexibility, scalability, adaptability, and extensibility to future systems. These features are not unique to statistical systems; see MIT (2004) and Rozanski and Woods (2005) for related comments in a broader context.

1.2. Improving the Balance of Total Survey Quality, Cost and Risk in the Context of Stakeholder Utility

Because statistical system architecture has developed to improve the balance of quality, cost, and risk in production systems, it is useful to consider this architecture in the context of total survey design and stakeholder utility. For general background on total survey design, see, for example Biemer (2010), Groves and Lyberg (2010) and references cited therein. For the current development, we note that a "total survey design" approach can provide a useful conceptual framework in which to develop, implement, and evaluate a wide range of methodological features, including work with frames, sampling, instruments, fieldwork, estimation, and inference. In addition, schematic extensions of the total survey design approach may help to provide a stronger and more systematic basis for the integration of methodology with production systems and survey management. This would involve three long-term goals:

(1) Provide a framework in which to explore the impact that specific features of the system design may have on the balance among multiple performance criteria, including data quality, cost, risk, and stakeholder utility. Each of these performance criteria are multidimensional, and require evaluation over the full lifecycle of a production system, including initial design and development, implementation, continuing operations, maintenance, and periodic upgrades. This exploration may involve direct effects of the system design as such, as well as interactions between the systems design and relevant features of the methodological and administrative design.
(2) Use the framework in (1) to guide practical decisions related to resource allocation and governance processes for methodological, systems, and administrative design.
(3) Use the results from (1) and (2) to identify specific areas of research that are most likely to have a substantial impact on work at the interface of methodological, systems, and administrative design.

The current article will focus on schematic extensions related to goal (1), while goals (2) and (3) will be considered in future work. Consider a general survey design characterized by a vector X, where the term "design" includes both formal design decisions and implementation of those decisions. Components of X generally would include all features of standard survey procedures, for example: development and refinement of frames and related administrative record systems; sample designs; collection methods; review, edit, imputation, and estimation methods; and procedures for dissemination of the resulting estimates. Additional components of X include features of the computational systems used to implement the abovementioned survey procedures, and features of administrative processes that may have an effect on survey quality, cost, and risk.

In the development of this article, it will be useful to distinguish among factors associated primarily with methodological (M), systems (S), and administrative (A) features of the design, and we will partition the design vector $X = (X_M, X_S, X_A)$ and related model parameter vectors accordingly. In addition, performance of our survey design may be affected by uncontrolled factors Z. These factors may include general features of the statistical environment (e.g., quality of frame information, accessibility of sample units through prospective collection modes, and the willingness of specific subpopulations to cooperate with survey requests), availability of resources, operational constraints, and changes in stakeholder needs for data. The factors Z also include process data and other forms of metadata and paradata (e.g., data on contact histories, costs, and effort expended at the unit level of effort, and incomplete-data patterns) as discussed in, for example, Couper (1998) and Kreuter et al. (2010).

We emphasize that each of X_M, X_S and X_A may depend heavily on the statistical production system components of the architecture. Also, the sensitivity of these components to changes in Z may depend on the architecture.

We can characterize the performance of the resulting survey design in terms of its quality, cost, risk, and stakeholder utility. In most practical cases, each of these performance measures will be multidimensional. First, let Q represent standard measures of data quality. This might include the components of a standard total survey error model, as well as more qualitative properties like timeliness and relevance, as reviewed by Brackstone (1999) and others. The quality measures may also include criteria developed originally in the computer science and information systems literature, for example, use of generally accepted systems design features, transparency of system architecture and code, or performance of the system with specified test cases.

Second, let C represent the costs associated with the development, implementation, maintenance and operation of the procedures; this includes costs for field operations, systems work, dissemination and methodology. One may consider schematic models for quality and cost, for example,

$$Q = g_Q(X, Z, \beta_Q) \tag{1.2.1}$$

and

$$C = g_C(X, Z, \beta_C) \tag{1.2.2}$$

where $g_Q(\cdot)$ and $g_C(\cdot)$ have (nominally) known functional forms, and β_Q and β_C are parameter vectors. In many cases, some or all components of β_Q and β_C are unknown.

Third, note that standard approaches to survey quality have tended to focus on issues that arise from cumulative effects of a large number of random or quasi-random events, such as separate household-level decisions on whether to respond to a survey request. However, the performance of a statistical production system can be seriously degraded by a small number of discrete events. For example, in many large-scale surveys, managers devote considerable effort to prevention of catastrophic degradation in one or more dimensions of quality. These efforts may include reduction of the probability of a given failure, as well as development of system features that will ameliorate the effects of failures that do take place. As a second example, Linacre (2011) and others have noted that in the development of survey systems, there are substantial risks that the development project might be aborted for a variety of reasons. For example, in many cases, termination took place when it became clear that initial assessments of cost or quality effects were unrealistic.

In principle, one could expand the definition of "survey quality" to include the abovementioned problems. However, since previous literature on survey quality has generally not included these issues, we will instead use a separate definition of *survey risk* R as the combined impact of a substantial and discrete failure in one or more components of a survey process. This would include degradation in survey quality, as well as failures in database management; in system development, testing and maintenance; in confidentiality protection for disseminated data; or in standard electronic data security. Characterization of these components of risk generally will include the severity of the failure, the duration and trajectory of recovery, effects of resource re-allocation arising from the failure, and the related impact on institutional credibility and on perceived stakeholder utility. Thus, for this discussion, the term "risk" incorporates the probability of a specific failure, its direct and indirect costs, and its broader impact.

As in the case of quality and cost, the survey risk R generally will be multidimensional, and we can consider a model

$$R = g_R(X, Z, \beta_R) \tag{1.2.3}$$

where $g_R(\cdot)$ is a function of general form and β_R is a vector of unknown parameters.

Fourth, let U represent the perceived stakeholder utility of the resulting survey design as implemented, and consider a model

$$U = g_U(Q, C, R, X, Z) \tag{1.2.4}$$

intended to link perceived utility with measures of quality, cost, and risk. In some relatively simple settings, R is perceived to be negligible, $g_U(\cdot)$ is the identity function and $U = Q$. However, in many cases, perceived utility is influenced by components of the design or environmental factors that may not be captured directly through standard quality measures Q. Furthermore, perceptions of utility may be influenced by components of risk and cost. For example, the risk that data releases will not occur on schedule may reduce the perceived value of these releases. Similarly, if cost structures are unpredictable from one year to the next, then data availability in a future year may be problematic, which may in turn reduce the perceived utility of the proposed production system.

In addition, note that Model (1.2.4) includes Q, C, and R (which are functions of X and Z), and also incorporates X and Z directly. For example, let X_1 be the design factor that determines the level of geographical granularity in the publication; this may have a direct effect on the perceived utility of the data release. On the other hand, let X_2 be the design factor that determines sample allocation across strata; this factor may affect U only through the components of Q related to estimator variance.

Fifth, note that each of the performance measures Q, C, R and U will have distinct subvectors associated, respectively, with development and implementation of the design, as well as with ongoing production and maintenance. To simplify the exposition, we will use "performance" as an umbrella term for all of our quality, cost, risk, and utility measures at any point in the cycle of design, implementation, operations, and maintenance, and will use

$$P = g_P(X, Z, \beta_P) \tag{1.2.5}$$

to define the resulting "performance surface," which is somewhat analogous to a multivariate "response surface" considered in industrial quality control. In addition, we emphasize that there are often complex multivariate relationships among Q, C, R and U, and one could expand Model (1.2.5) to reflect these relationships.

Finally, in keeping with the term "schematic" we emphasize that Models (1.2.1)–(1.2.5) are intended primarily to provide some structure for discussion of the complex trade-offs among cost, quality, risk, and perceived utility encountered in large-scale statistical work, and should not be interpreted in an excessively literal or reductionist form. For example, some components of P may be measured with a reasonable level of precision, while other important components may allow only very general qualitative assessments. Similarly, for some components of P, relationships with factors X and Z may be reasonably approximated by, for example, a linear or nonlinear regression model, while in other cases the relationships are highly fluid and not reasonably characterized or estimated through standard statistical models with available data.

1.3. *The Combined Effects of Methodological, Systems, and Administrative Components of the Design Vector X*

To address goals (1) through to (3) presented in Section 1.2, consider a relatively simple linear form of the performance model:

$$P = g_P(X, Z, \beta_P) = \beta_{P0} + \beta_{PM}X_M + \beta_{PS}X_S + \beta_{PA}X_A + (interactions) + e_P \qquad (1.3.1)$$

Under Model (1.3.1), β_P contains the intercept vector β_{P0}, the main-effect coefficient vectors $\beta_{PM,}$ $\beta_{PS,}$ β_{PA}, and β_{PZ} and the coefficient vectors for the two-factor and higher-order interactions as needed. In addition, the error term e_P is assumed to have a mean of zero and a variance-covariance matrix V_{eP} that may depend on the design variables X and the environmental and process variables Z.

Eltinge (2012) and Eltinge and Phipps (2009) explored variants of Model (1.3.1) for general survey design work. The current article focuses primary attention on the effect that the system design vector X_S has on the mean and variance functions of the performance measures P. Specifically, in the use of Model (1.3.1), three topics of special interest are:

(a) System features X_S that allow the measurement or (partial) control of specified components of P. Examples include features that facilitate the collection and analysis of data from a contact history instrument (CHI); computer-assisted recording (CARI); cost accounting records; audit trails produced by microdata edit and review processes; or disaggregated reports of interviewer activity linked with survey responses.

(b) System features X_S that impose constraints on some components of the methodological and administrative design vectors ($X_{M,}$ X_A), or on the use of information regarding environmental factors Z. For example, use of certain standardized system features may preclude some types of local optimization of a methodological design; or may limit the amount of process data that one may collect on specific measures of cost, quality, or risk. Moreover, system constraints that do not allow for poststratification or response-propensity adjustments would restrict the use of such environmental information in Z for the improvement of survey weights.

(c) Direct effects of X_S on P, through the main-effect coefficient β_{PS} and related interaction coefficients. For example, some system features may allow for improved editing of collected data; timelier follow-up for nonrespondents; more flexible on-demand production of estimates for special subpopulations; or real-time monitoring of estimates based on data collected to date, as considered in some forms of responsive design. In some cases, one may also consider the effect of X_S on V_{eP}, in keeping with the study of dispersion effects in the response surface methodology literature (e.g., Brennerman and Nair 2001, Bisgaard and Pinho 2003 and references cited therein). However, dispersion effects appear to have received relatively little attention in the survey and systems design literature to date.

The remainder of this article explores issues (a), (b) and (c) through the following steps. Section 2 considers the use of the performance model (1.3.1) to inform design decisions. Section 3 explores the relationship between the general model (1.3.1) and more specific models developed for Total Statistical Quality and Adaptive Total Design. Section 4 links the model, and related design decisions, with the Generic Statistical Business Process Model (GSBPM) and the Generic Statistical Information Model (GSIM). Section 5 outlines related issues with governance processes that link the managerial design features X_A with the system design features X_S.

2. Use of Performance Models to Inform Design Decisions

2.1. General Questions for Exploration of Design Decisions

Ideally, one would have extensive information on all relevant aspects of the performance model (1.3.1), including identification of the applicable predictors X and Z (incorporating both main effects and interactions); sufficiently precise estimates of the coefficients β_P; and the values of R^2 and other goodness-of-fit measures. Under such conditions, selection of a preferred design X (including the systems design features X_S) is arguably a straightforward exercise in optimization or satisficing, based on the specified multivariate performance criteria.

In most practical cases, however, currently available empirical information falls far short of the abovementioned ideal. Consequently, it is important to apply standard concepts from social measurement and response surface methodology to describe the more limited types of design information that may be feasible to obtain, and that may be useful in practical work with the systems design features X_S. The following questions are of particular interest:

(A) What practical design decisions are under consideration? Examples include:

(A.1) Design of a survey that is mostly or entirely new. For this case, one would seek to understand the extent to which previous experience with somewhat similar surveys may offer partial insights into the performance surface for the proposed new survey. Beyond that limited information, evaluation of the performance surface will depend primarily on new empirical work, such as small-scale tests of system components and of the interfaces of those system components with specific methodological or managerial design components.

(A.2) Comprehensive revision of one or more survey design features, while leaving other design features largely unchanged. Here, it will be important to identify the extent to which design components that are subject to change will have substantial interactions with other factors in the performance surface (1.3.1). If the changed components influence the performance surface only through main effects, then it may be relatively simple to assess the likely performance of the survey under the proposed new design. On the other hand, if the changed components enter Model (1.3.1) through complex interactions, then evaluation of performance under the new design may require extensive amounts of additional empirical work.

(A.3) Incremental revision of one or more design features. In this case, detailed empirical information on the performance surface (1.3.1) under the previous design may provide important insights into likely performance under the new design. This final case is somewhat analogous to "evolutionary operation" approaches considered in industrial response surface methodology. See, for example, Box (1957) and Box and Draper (1969).

(B) Out of the many potential performance characteristics reviewed in Section 1, which features are of primary interest to the survey organization? Are there practical and objective methods available to measure these performance characteristics? If not, what methodological, systems or management work should one carry out to develop and implement such measures?

(C) Within the context defined by parts (A) and (B):

(C.1) What information is currently available regarding Model (1.3.1), including its functional form, dominant main effects and interactions, and parameters β_P and V_{eP}?

(C.2) Does the information in (C.1) provide sufficient information to guide the practical decisions identified in (A)? The definition of "sufficient information" will depend heavily on the decisions of primary interest in (A). For example, for the extensive redesign work in cases (A.1) or (A.2), it may suffice to have very rough order-of-magnitude indications of the improvements in quality or cost likely to follow from the proposed changes in X. Conversely, one may consider incremental revisions considered in Case (A.3) as part of a modest effort to obtain moderate improvements in quality or efficiency; and the corresponding definition of "sufficient information" may depend primarily on the variances of available estimators of the coefficient vector β_P. Similar comments apply to cost, quality and risk measures related to the system development process itself.

(C.3) If the answer to (C.2) is no, what additional information is needed, and what methods should be used to obtain that information (e.g., formal experiments, pilot tests, or comparisons with industry benchmarks)?

(C.4) During development, implementation, operation and maintenance of a designed production system, what are the ways in which available information will be used to monitor performance and identify needed mid-course adjustments? This information may include model parameter data from (C.1)–(C.3), the performance measures from (B), and related paradata. In keeping with the cautionary note by Linacre (2011) cited above, it is especially important to use relevant measures of cost, quality and risk incurred during the process of system development and implementation, as well as performance measures that apply to operation and maintenance of the production system after implementation.

(D) Are there external constraints that have an important effect on the design project? These constraints may include standard methodological and managerial constraints, as well as imposition of limits on certain system design features. For example, senior management may mandate the use of specified security software, increased use of standardized system features, or reduced funding available for specified types of design work.

2.2. *Questions Arising from Dynamic Features of Quality, Cost, Risk, and Utility*

One may use Model (1.3.1) to characterize a survey process that is static, in the sense that (after completion of design implementation) the coefficients β_P are essentially fixed, the design vector X is generally not subject to change, and the distribution of the paradata Z is relatively stable. For this reason, some persons who specialize in one area of methodology, management, or systems may tend to perceive the other areas as readily characterized by a fixed set of "requirements."

In many practical settings, however, the schematic models presented in Section 1.2 are fundamentally dynamic in nature. For example, some environmental factors Z may vary over time, and thus may have an effect on quality, cost, risk, and stakeholder utility. To take one methodological case, changes in cell telephone and landline telephone use over the past two decades have had a substantial effect on the quality characteristics of standard random digit dialing surveys. In addition, during the recent recession, many stakeholders had increased interest in unemployment rates and other labor-force characteristics, which in turn led to changes in the linkage between perceived utility and standard quality measures like accuracy and timeliness.

Similarly, both survey methodology and systems work are dynamic in nature, so the schematic models for quality, cost, risk, and utility may change over time. Some relatively simple examples of changes include the maturing of specific methodological features (e.g., more sophisticated use of incentives or nonrespondent-conversion techniques); improved testing and robustness of complex skip patterns in CATI or CAPI instruments; or amortization of costs related to methodology, training or systems development and implementation.

Dynamics in the development and maturation of methodology imply that system architecture work often must go beyond specification and satisfaction of static "system requirements." Conversely, dynamics in systems architecture, and in the needs of data users, imply that methodologists must be open to considering a wide range of performance criteria, some of which may differ substantially from traditional statistical performance criteria like bias and mean squared error.

Consequently, it may be important to supplement questions in areas (A) through (D) with additional questions that are linked directly with the dynamics of the survey process. Some examples include:

(E.1) In keeping with a suggestion by Hidiroglou (2011, personal communication), what are appropriate practical criteria to use to determine whether a given methodological or managerial component has reached a level of "maturity" that warrants incorporation into standard production systems?

(E.2) To what extent, and in what ways, should one try to build flexibility into standardized systems to increase the likelihood that they will accommodate further methodological refinements or managerial changes at a relatively low cost?

2.3. Approximations Arising from Standardization Processes

As noted in the preceding sections, work with system architecture often leads to discussion of standardization of some or all of the components of a survey process across surveys. Potential benefits of standardization include improvement in some components of data quality (especially those related to failure to execute design features as specified), and reduction of certain components of cost and risk. On the other hand, in some cases standardization may prevent a given survey from making optimal use of survey-specific information on, for example, characteristics of the underlying target population. In that sense, an excessively rigid application of standardization may lead to degradation of some components of survey performance. Similar issues have been noted for standardization in some areas far removed from survey work. For example, Hall and Johnson (2009) discuss distinctions between environments in which a process may be improved through the application of "art" or "science" respectively. Extending their general ideas to the statistical performance model (1.3.1), one could suggest that extensive use of standardization may be warranted when one can account for most of the variability of cost, quality, risk and utility through some simple predictors *X* that correspond to highly standardized design features. Conversely, standardization may be problematic when much of the variability of cost, quality, risk and utility are attributable to the uncontrolled environmental factors *Z,* unobserved remainder terms e_P, or design components *X* that correspond to highly customized design decisions.

3. Relationship With Models for Total Statistical Quality and Adaptive Total Design

Sections 1 and 2 were developed in the context of general performance measures *P*. To date, the survey literature has devoted a large amount of attention to some quality measures *Q*, some attention to cost structure (e.g., Groves 1989), and relatively little attention to survey risk *R* and stakeholder utility *U*. Consequently, this section reviews some previously developed strategies for continually improving the quality of system outputs in real time as the outputs are being generated. In previous literature, this approach has been given several different labels, including active management (Laflamme et al. 2008), responsive design (Groves and Heeringa 2006) and adaptive design (Schouten et al. 2011). It may be of interest to develop similar approaches for *C*, *R* and *U* in future work. As with many cases involving high-dimensional objective functions, a focus on specific subcomponents can lead to suboptimization, but nonetheless may be necessary for tractability.

For the current discussion, we will focus on a strategy known as Adaptive Total Design (ATD), as presented in Biemer (2010). ATD aims to reduce the risks of poor data quality in survey processes by using concepts embodied in the total survey error paradigm. In the notation defined in Section 1, ATD seeks to maximize *Q* under a constraint on *C* by selecting design features *X* so that *Q(X,Z)* is maximized subject to the constraint *C(X,Z)* $< C_T$ where C_T is the data collection budget.

Alternatively, the ATD may seek to minimize the total risk, *R(X,Z),* subject to the same budget constraint. At the present time, ATD has only been applied for the data collection process. See, for example, Carly-Baxter et al. (2011) and Cunningham et al. (2012). In addition, some important features of ATD have been implemented in other data collection work. For example, Calinescua et al. (2013) apply some elements of ATD to control nonresponse and measurement error in the initial design, but their approach does not incorporate quality monitoring and improvement during data collection. Wagner (2008) uses adaptive design to reduce the effect of nonresponse on the estimates during data collection; however, his work does not attempt to control other types of nonsampling errors, such as measurement error or frame error. Further generalizations to other components of the GSBPM framework are conceivable and will be discussed subsequently.

To illustrate the principles of ATD for real-time, quality improvement we begin with a preliminary assessment of *Z*, say Z_0, and available budget C_T. We choose an initial design (or, as in the current case, a data collection approach) X_0 which we believe will produce the optimal quality level $Q*$. However, typically, the quality level produced by X_0 will not be optimal because either g_Q and/or g_C are usually misspecified in the initial design or the initial assessment of Z was not current or otherwise did not describe the true, current process environment. Suppose that, during the implementation of the process, new information on Z, say Z_1, is obtained. Based upon Z_1, the initial design X_0 may no longer be optimal; that is, the quality level that can be achieved by X_1 is $Q_1 < Q*$. In order to achieve a quality level that better approximates $Q*$, the preliminary design must be altered to some extent, say, to X_1 while still ensuring that the budget, C_T, is not exceeded. As the process continues to completion, there may be new assessments of Z, say, Z_2, Z_3,. . . with consequential design alterations, X_2, X_3, ...

ATD relies on one's ability to obtain updated information on Z during production and perhaps better specifications of g_Q and/or g_C so that $Q(X,Z)$ can be re-estimated, leading to a re-optimization of the design. For example, in the data collection setting, information on interviewer field performance may be continually updated as the interviewers work on their assignments. This information may relate to the cost of interviewing by area, difficulties in contacting and interviewing certain types of sample units, sample unit cooperation as a function of the unit characteristics, and so on. This information and other paradata streaming in from the field constitute Z_k for $k = 1, 2, \ldots$.

As the design is altered during implementation to achieve quality levels that approach $Q*$, the budget must be continually assessed requiring precise estimates of the cost to completion at each stage. For example, at the start of data collection, the full budget C_T was available. However, by the time that Z_1 is observed, only C_1 of the original budget remains; that is, $C_T - C_1$ has been spent by the time Z_1 is observed. Therefore, in determining what interventions $X_1 - X_0$ are needed to achieve $Q*$ under Z_1, the currently available budget, C_1, must not be exceeded. Thus, the shift from X_0 to X_1 is made that will produce a final level of quality Q_1 where $Q_0 < Q_1 \leq Q*$ where the costs of X_1 are constrained by C_1. Note that the difference $X_1 - X_0$, referred to above as design "interventions," are changes in the original design necessary to achieve the desired level of quality. This process of iteratively re-evaluating Z (i.e., Z_1, Z_2, . . .) and applying interventions to achieve designs X_1, X_2, X_3, . . . such that $Q_1 < Q_2 < Q_3 < \cdots \leq Q*$ while holding total costs fixed at C_T is referred to as Adaptive Total Design (Biemer 2010).

To summarize, ATD begins with an initial design X_0 which is optimal given Z_0, and then seeks to find a sequence of designs $X_1, X_2, \ldots, X_k$ based upon the updated information Z_1, Z_2, . . . and updated functions g_Q and g_C such that $Q(X_k, Z_k) = Q*_k$, the maximum quality that can be achieved for a total cost of C_T. ATD can be thought of as a form of Evolutionary Operation (EVOP; e.g., Box 1957; Box and Draper 1969 and Hahn 1982) for data collections where the goal is to find the maximum of the performance surface under cost constraints.

Some of the implications of ATD can be summarized as follows:

a. Paradata feeds the EVOP process by providing updated information on Z and the specifications of g_Q and g_C.
b. Therefore, the goal of collecting paradata should be to:

   i. best inform those aspects of Z, g_Q, and g_C that are most predictive of Q and C and
   ii. best capture the dynamic nature of the data collection process.

c. It is important to monitor these paradata in order to maximize data quality because the difference $Q_k - Q_1$ (i.e., the quality improvement after k ATD cycles) may be quite large.
d. When the budget is reduced, the reduced budget, say $C_{T1} < C_T$, is entered into the ATD optimization and the design X is reoptimized. This usually results in a lower level of quality that can be achieved.

4. Possible Integration With the Generic Statistical Business Process Model (GSBPM) and the Generic Statistical Information Model (GSIM)

Sections 2 and 3 explored the ideas of Section 1 in the context of general performance models and Adaptive Total Design, respectively. Detailed conceptual and empirical expansion of Section 1 also would require a detailed characterization of survey processes. Over the past decade, several organizations have developed forms of this characterization. See Dunnet (2007), Renssen and Van Delden (2008), Gloersen and Saeboe (2009), and Pink et al. (2010). An effort to standardize these approaches and related terminology through a Common Metadata Framework has led to a Generic Statistical Business Process Model (GSBPM; Vale 2009) based largely on Statistics New Zealand's business process model.

The GSBPM has quite rapidly become a standard instrument to communicate *between* organizations. This outcome arose from the resemblance of GSBPM to already existing descriptive models, its successful union of terminology for higher level statistical production processes, and good communication and anchoring. The GSBPM was identified by the High-Level Group for Strategic Developments in Business Architecture in Statistics (HLG-BAS) as a key standard for modernizing official statistics (UNECE 2011 and UNECE 2012a).

It would be premature to forecast specific impacts of the GSBPM, but its shared process terminology has the potential to accelerate and improve development of common areas. For example, it has already been used to pinpoint collaboration areas and to a certain extent it helped in the harmonization of architectures (see Vaccari 2009). In addition, some authors have suggested that the GSBPM also could be a framework for process quality assessment and improvement (Todorov 2011) and that it will facilitate the sharing of statistical software and components (Hamilton and Tam 2012a, 2012b).

An important complement to the GSBPM is the Generic Statistical Information Model (GSIM). GSIM is described as "a reference framework of internationally agreed definitions, attributes and relationships that describe the pieces of information that are used in the production of official statistics (information objects). This framework enables generic descriptions of the definition, management and use of data and metadata throughout the statistical production process" (UNECE 2012b).

In particular, GSIM aims to define and describe in a harmonized way the *information objects* that flow through and control the statistical production processes. The resulting common framework may help development of shared production systems by improving communication between distinct disciplines, for example, subject matter statisticians, IT specialists and methodologists. GSIM is also aligned to standards such as DDI (Data Documentation Initiative; http://www.ddialliance.org/) and SDMX (Statistical Data and Metadata eXchange; http://sdmx.org/), but it does not depend on them.

In the future, the GSBPM and GSIM (or similar frameworks) could be valuable for exploration of the schematic approach outlined in Section 1, and related topics considered in this special issue of the Journal of Official Statistics. Two cross-cutting areas may warrant special attention. First, GSBPM (or similar national models) could lead to standardized and perhaps highly modular production systems and architectures, and GSIM could support standardized handling of information objects in a complex production system. Stated in terms of Models (1.2.1)–(1.2.5), the resulting modularity could have an effect on system performance through both the X_S and Z factors. Some effects may be related to X_S because modularity (more subsystems/components) will require system design with higher attention directed toward the system's life cycle factors. If a system fails when one of its components fails, then increased standardization and reuse of components will increase the importance of managing the full spectrum of risks. This includes risks associated with individual components, integration of these components, and performance of the overall production chain. In addition, if standards and harmonization of information objects lead to more system components that share data, then the resulting production systems may have performance that is less subject to uncontrolled environmental factors Z. An arguably desirable development would be that some factors that today are observable but not controlled (i.e., Z terms) can be replaced by controllable design factors X_M or X_S. One such example could be automation of certain parts of a production system. Hamilton and Tam (2012a, 2012b) and Borowik et al. (2012) provide thoughts about possible future modular production systems where attention to these matters is needed.

Second, one could consider expansions of the GSBPM that would incorporate in-depth collection and use of information on costs and data quality throughout the process chain.

Although quality management is present in the model as an overarching process, it is not reflected specifically in the model phases or subprocesses. For example, as noted above, an important part of the design work is the design of ATD monitoring and improvement components that run simultaneously with, and gather data from, the processes. These components are critical for updating Z and respecifying X to optimize Q in real time; and are implemented later, during data collection. Likewise, our ability to conduct post-survey data quality evaluations requires that components are in place to capture the relevant data on survey errors that aid these evaluations. In addition, systems for quality control are key components of work with initial development, as well as data collection and processing. Quality control components need to be developed and implemented to assure both product and process quality. Finally, documentation on data quality is a critical component of data dissemination and archiving. These are just a few areas of the GSBPM where quality management could be emphasized by providing specific references to subprocesses that address ATD-related improvements to data quality. In this regard, the current version of the GSBPM seems to emphasize statistical standards (i.e., standards related to data production). However, it provides relatively little coverage of quality standards (i.e., standards related to quality improvement) and of quality functions like the measures Q

considered in Section 1 above. The European Code of Practice (available from http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code = KS-32-11-955), which focuses primarily on developing, producing and disseminating European statistics, provides some examples of statistical standards. An example of a quality standard is Eurostat's Quality Assurance Framework (available from http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/QAF_version_1.0_EN.pdf), which focuses primarily on quality assurance and quality management.

5. Governance Processes and Management of Stakeholder Expectations

The development, implementation, testing, operation and maintenance of survey production systems require cooperative work by several distinct groups within a survey organization. These groups tend to have different skill sets, institutional cultures and incentive structures. Consequently, governance processes for architecture and systems work can involve a complex set of factors. In a formal sense, these factors can be viewed as some of the components of the "management design" vector X_M introduced in Section 1. This section presents some of these management factors. Subsection 5.1 outlines some general features of a governance process. Section 5.2 reviews some related features of institutional and professional environments. Throughout this section, we use the term "management" as the omnibus term for all relevant management activities.

5.1. Features of a Governance Process

Roles and responsibilities of management. These include the following:

- Resource allocation, including (a) initial investments; (b) internal pricing for use of production systems (e.g., a free good; partially subsidized use; or full assignment of costs to users; and procedures for amortization of costs, based on anticipated longevity of a given system or component); (c) mid-course adjustments and project termination.
- Acquisition and evaluation of relevant and actionable information regarding the quality, cost and risk factors that are most directly relevant to decisions on all aspects of the survey process. This generally will require management personnel to have extensive training and experience in the relevant technical areas.
- Resolution of disagreements among stakeholders, especially in relation to trade-offs between distinct components of cost, quality and risk. This will require senior personnel to have broad skills for leadership and consensus-building in environments for which ambiguity and incomplete information may be quite prominent.
- Communication and implementation of decisions. Depending on the specific decisions, communication may include a wide range of internal and external stakeholders. Implementation work includes all of the steps required to convert an abstract decision into concrete results, and may be especially challenging when management has limited information on the dominant features of the performance models (1.2.5).
- Personnel management. This includes determination of skills and experience required for specific tasks; full-time or part-time assignment of specific personnel to

these tasks; integration of these assignments with general career paths in the survey organization; and evaluation of individual and group performance for these tasks.

Mechanisms for management. This includes both control of X and the methods used to control specific components of X. Examples of management mechanisms include standard direction through a supervisory chain and management through a committee structure. For committee-based management, clearly defined decision processes may be especially important, and may include decisions by formal committee vote, decisions by formation of a consensus, and decisions by a line manager based on advice from the committee.

Feedback loops for management of specific design components. This includes feedback loops to provide information on (1) adequacy or limitations of models for C, Q, R and U; (2) operational risk (e.g., identification of cases in which there are substantial deviations from the nominal design during implementation); (3) environmental factors Z with large coefficients for the main effects on Z, or for the interactions between X and Z.

Incentives for individuals or groups. Successful revisions of system architecture generally require that the relevant individuals and groups have an incentive structure that is aligned clearly with the survey organization's need to improve its overall balance of cost, quality and risk, and to overcome institutional inertia that may stand in the way of that improvement. Examples of implicit or explicit incentives include: direct mandates for standardization and systems use; subsidized or free access to systems developed under the revised architecture; allowing the participating groups to re-invest or otherwise control some of the savings obtained through the revised architecture; or linkage of participation with individual promotions or other employment-related benefits. Mechanisms for implementation of these incentives include standard personnel management processes; internal competitions for resources used in systems development, testing and implementation; and allowed flexibility for customization.

5.2. *Relevant Features of the Institutional and Professional Environment*

In addition to the general features outlined above, the success of a governance process may depend on several qualitative features of the institutional and professional environment, and the extent to which those features are coherent with the performance characteristics of the proposed system.

Expectations regarding objective and measurable criteria for performance. Under a (generally unattainable) ideal, this would include all relevant dimensions of the C, Q, R and U vectors. More realistically, this would include dimensions of these vectors that are considered especially important for key stakeholders.

Professional norms and standards for transparency, credibility, intellectual property and management confidentiality. These norms and standards appear to vary substantially across different survey organizations, and across different professional groups within a survey organization. For example, among professionals in mathematical statistics and methodology, professional credibility generally depends on a very high degree of transparency regarding methods used and the strengths and limitations of those methods. Intellectual property rights generally center on publication copyrights and acknowledgement of priority, rather than formal patent or property rights as such, and confidentiality restrictions (with respect to management) generally are limited to certain business processes (e.g., bid pricing).

External benchmarking of system features and performance measures. Some organizations expect that most features of a survey process (including the general architecture, methodology, fieldwork, production systems and management processes) will be rigorously compared with relevant external benchmark standards. Other organizations carry out little or no external benchmarking, in some cases due to perceptions that their processes are too specialized to allow for meaningful external comparisons.

Use of objective external peer review processes. In parallel with the discussion on external benchmarking, some survey organizations make extensive use of external peer review processes, while others do not. For survey groups that do use external peer review processes, important issues include the selection and management of the peer review group (e.g., strong standards for expertise and credibility, and procedures to prevent conflicts of interest); the degree to which review recommendations are binding on the survey organization; and specific resource allocation decisions and institutional incentives that are linked with review recommendations. In addition, the effectiveness of an external peer review process will depend heavily on the extent to which the members of the review group have a carefully nuanced understanding of the relevant issues, based on their skills and experience, as well as sufficient access to survey-specific information. Architecture based on extensive sharing of system components may provide a degree of internal peer review for some system components, but is not a substitute for systematic external reviews. See Lyberg (2012) for in-depth discussion of external peer review processes for survey organizations.

6. Closing Remarks

Work with architecture for a statistical production system involves a complex balance among a wide range of performance measures that include cost, quality, risk, and stakeholder utility. Each of these measures may be influenced by design decisions regarding methodological, systems, and management factors, as well as other factors that are not subject to control. Models (1.2.1) to (1.2.5) provide a schematic framework through which one may discuss performance measures and design decisions for statistical production work.

The preceding five articles in this special issue of Journal of Official Statistics provide some snapshots of the current state of work at the interface of methodology with the architecture of statistical production systems. These articles report on recent progress at the abovementioned interface, and also identify areas in which additional work will be warranted. We believe that four general topics will be especially important: more systematic empirical assessment of factors that affect cost, quality, risk, and stakeholder utility; use of these empirical results to expand previously developed approaches to total survey error and total statistical design; integration of the schematic models (1.2.1) to (1.2.5) with the general process descriptions provided by the GSBPM and the GSIM; and careful development and implementation of governance processes that are tuned appropriately for the architecture of statistical production systems. We hope that this special issue will contribute to a more systematic study of these topics, and look forward with enthusiasm to further developments in this field.

7. References

Australian Bureau of Statistics (2012). International Collaboration Effort – Statistical Network Progress Report. Paper presented by Brian Pink at the meeting of Directors General of participating countries Australia, New Zealand, Norway, Sweden, Canada, UK in New York, February 2012.

Biemer, P.P. (2010). Total Survey Error: Design, Implementation and Evaluation. Public Opinion Quarterly, 74, 817–848.

Bisgaard, S. and Pinho, A. (2003). Follow-Up Experiments to Verify Dispersion Effects: Taguchi's Welding Experiment. Quality Engineering, 16, 335–343.

Borowik, J., Henden, M., and Fraser, B. (2012). Riding the Big Data Wave to Streamline Acquiring and Managing Data. Meeting on the Management of Statistical Information Systems (MSIS 2012). Washington, DC, 21–23 May 2012.

Box, George E.P. (1957). Evolutionary Operation: A Method for Increasing Industrial Productivity, Journal of the Royal Statistical Society. Series C (Applied Statistics). Vol. 6, No. 2 (June 1957), 81–101.

Box, G.E.P. and Draper, N.R. (1969). Evolutionary Operation. New York: Wiley.

Braaksma, B. (2009). Redesigning a Statistical Institution: the Dutch Case. In Proceedings of Modernisation of Statistics Production. International conference organised by Statistics Sweden, Stockholm, 2–4 November 2009. Available at: http://www.scb.se/Grupp/Produkter_Tjanster/Kurser/ModernisationWorkshop/final_papers/H_2_business_models_Braaksma_final.pdf. (accessed February 12, 2013).

Brackstone, G. (1999). Managing Data Quality in a Statistical Agency. Survey Methodology, 25, 139–149.

Brenneman, W.A. and Nair, V.N. (2001). Methods for Identifying Dispersion Effects in Unreplicated Factorial Experiments: A Critical Analysis and Proposed Strategies. Technometrics, 43, 388–405.

Calinescua, M., Bhulaia, S., and Schouten, B. (2013). Optimal Resource Allocation in Survey Designs. European Journal of Operational Research, 21, 115–121.

Camstra, A. and Renssen, R. (2011). Standard Process Steps Based on Standard Methods as a Part of the Business Architecture. 58th Congress of the International Statistical Institute. Dublin, 21–26 August 2011.

Carly-Baxter, L., Mitchell, S., and Peytchev, A. (2011). Developing and Implementing Adaptive Total Design. Presented at the 2011 International Field Directors and Technologies Conference (IFD&TC), Scottsdale, AZ. Available at: http://www.ifdtc.org/PC2011/presentation_2011_files/3A-Lisa%20Carley-Baxter.pdf. (accessed February 12, 2013).

Chief Information Officers Council CIOC (1999). Federal Enterprise Architecture Framework. Version 1.1, Chief Information Officers Council, September 1999.

Chief Information Officers Council CIOC (2001). A Practical Guide to Federal Enterprise Architecture. Version 1.0, Chief Information Officers Council, February 2001.

Couper, M. (1998). Measuring Survey Quality in a CASIC Environment. In: Proceedings of the Section on Survey Research Methods, American Statistical Association.

Cunningham-Hunter, D., Mitchell, S., Carley-Baxter, L., and Keating, M. (2012). Using Paradata to Monitor Survey Production, Cost, and Quality within an Adaptive Total Design Framework. Paper presented at the 2012 Federal Conference on Survey Methodology, Washington, DC. Available at: http://www.fcsm.gov/12papers/Hunter_2012FCSM_V-C.pdf. (accessed February 12, 2013).

Dunnet, G. (2007). The BmTS: Creating a new business model for a national statistical office of the 21st century. Invited paper at the UNECE/Eurostat /OECD Meeting on the Management of Statistical Information Systems (MSIS). Geneva, 8–10 May 2007. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2007/mtg1/wp.11.e.pdf. (accessed February 13, 2013).

Eltinge, J.L. (2012). Doing More with Less: Balancing Survey Cost, Quality, Risk and Stakeholder Utility. Presentation at the Fourth International Conference on Establishment Surveys, Montreal, Canada.

Eltinge, J.L. and Phipps, P.A. (2009). Characterization, Evaluation and Management of Prospective Benefits, Costs and Risks in the Development of New Statistical Programs for Energy. Proceedings of the Section on Survey Research Methods, Joint Statistical Meetings.

European Commission (2009). Communication from the Commission to the European Parliament and the Council on the production method of EU statistics: a vision for the next decade, Document COM(2009)404. Available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri = COM:2009:0404:FIN:EN:PDF. (accessed February 12, 2013).

Field, S. (2009). From Red to Green: The role of Enterprise Architecture in the ONS Corporate IT Strategy. UNECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS 2009). Oslo, Norway, 18–20 May 2009.

Finselbach, H. (2011). Implementing Standardised Systems, Processes, Tools and Methods with the ONS Design Charter. 58th Congress of the International Statistical Institute. Dublin, 21–26 August 2011.

Gloersen, R. and Saeboe, V. (2009). Standardisation for Improvements in Statistics Norway. In Proceedings of Modernisation of Statistics Production International conference organised by Statistics Sweden, Stockholm, 2–4 November 2009. Available at: http://www.scb.se/Grupp/Produkter_Tjanster/Kurser/ModernisationWorkshop/final_papers/H_1_business_models_Gloersen.pdf. (accessed February 12, 2013).

Groves, R.M. (1989). Survey Errors and Survey Costs. New York: John Wiley & Sons.

Groves, R.M. and Heeringa, S. (2006). Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs. Journal of the Royal Statistical Society Series A: Statistics in Society, 169(Part 3), 439–457.

Groves, R.M. and Lyberg, L. (2010). Total Survey Error: Past, Present and Future. Public Opinion Quarterly, 74, 849–879.

Hahn, G.J. (1982). Statistical Assessment of a Process Change. Journal of Quality Technology, 14, 1–9.

Hall, J.M. and Johnson, M.E. (2009). When Should a Process be Art, not Science. Harvard Business Review, March 2009 issue, 58–64.

Hamilton, A. and Tam, S.-M. (2012a). Toward GSIM V1.0 as a Cornerstone for Common Reference Architecture. Meeting on the Management of Statistical Information Systems (MSIS 2012). Washington, DC, 21–23 May 2012.

Hamilton, A. and Tam, S.-M. (2012b). Platform for International Collaboration – Preliminary Thoughts From the ABS. Meeting on the Management of Statistical Information Systems (MSIS 2012). Washington, DC, 21–23 May 2012.

Jug, M. (2009). Uptake of Service Oriented Architecture in Statistics – Are We Really So Different? In Proceedings of Modernisation of Statistics Production International conference organised by Statistics Sweden, Stockholm, 2–4 November 2009. Available at: http://www.scb.se/Grupp/Produkter_Tjanster/Kurser/ModernisationWorkshop/final_papers/C_1_SOA_Jug_final.pdf. (accessed February 12, 2013).

Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T.M., Casas-Cordero, C., Lemay, M., Peytchev, A., Groves, R.M., and Raghunathan, T.E. (2010). Using Proxy Measures and Other Correlates of Survey Outcomes to Adjust for Nonresponse: Examples from Multiple Surveys. Journal of the Royal Statistical Society, Series A, 173, 389–407.

Laflamme, F., Pasture, T., Talon, J., Maydan, M., and Miller, A. (2008). Using Paradata to Actively Manage Data Collection, Proceedings of the ASA Survey Methods Research Section, Denver, Colorado.

Linacre, S. (2011). Standardising Methods and Tools: A Look Back Over 30-Plus Years of Experience. Paper presented at Statistics Canada Symposium 2011.

Lyberg, L. (2012). Survey Quality. Waksberg Memorial Lecture delivered at the Fourth International Conference on Establishment Surveys, 14 June 2012.

MIT (2004). Engineering Systems Division Architecture Committee: The Influence of Architecture in Engineering Systems. Paper presented at the Engineering Systems Symposium, MIT, Cambridge MA, USA, 29–31 March 2004. Available at: http://esd.mit.edu/symposium/monograph/default.htm. (accessed February 12, 2013).

Munoz-Lopez, J. (2012). Value of Systems Architecture for Statistical Organizations. Meeting on the Management of Statistical Information Systems (MSIS 2012). Washington, DC, 21–23 May 2012.

OMB (2009). Improving Agency Performance Using Information and Information Technology EA Assessment Framework v3.1, Office of Management and Budget. June 2009.

Penneck, S. (2009). The Office for National Statistics (ONS) Statistical Modernisation Programme: What Went Right? What Went Wrong? In Proceedings of Modernisation of Statistics Production, International conference organised by Statistics Sweden, Stockholm, 2–4 November 2009. Available at: http://www.scb.se/Grupp/Produkter_Tjanster/Kurser/ModernisationWorkshop/final_papers/D_1_management_Penneck_final.pdf. (accessed February 12, 2013).

Pink, B., Borowik, J., and Lee, G. (2010). The Case for an International Statistical Innovation Program –Transforming National and International Statistics Systems. Statistical Journal of the International Association for Official Statistics (IAOS), 26, Number 3–4. Available at: http://iospress.metapress.com/content/2h5764574t6318r4/fulltext.html. (accessed February 12, 2013).

Renssen, R. and Van Delden, A. (2008). Standardisation of Design and Production of Statistics; A Service Oriented Approach at Statistics Netherlands. Paper presented at the IAOS Congress Reshaping Official Statistics, Shanghai, October 2008.

Renssen, R., Wings, J., and Paulussen, R. (2009). Processes, Methods and Tools. Statistics Netherlands internal report. Available at: http://www.cbs.nl/NR/rdonlyres/8EFDD386-0FD9-4567-90BC-6ADAB5C67B3D/0/2009Processesmethodsandtoolsart.pdf. (accessed February 12, 2013.).

Rozanski, N. and Woods, E. (2005). Software Systems Architecture. New Jersey, United States of America: Addison Wesley.

Schouten, B., Calinescu, M., and Luiten, A. (2011). Optimizing Quality of Response Through Adaptive Survey Designs, Statistics Netherlands, Available at: http://www.cbs.nl/NR/rdonlyres/2D62BF4A-6783-4AC4-8E45-12EF20C6675C/0/2011x1018.pdf. (accessed February 12, 2013).

Sowa, J.F. and Zachman, J. (1992). Extending and Formalizing the Framework for Information Systems Architecture, IBM Systems Journal, 31.

Statistics Canada (2012). Corporate Business Plan – Statistics Canada – 2012–13 to 2014–15. Available at: http://www.statcan.gc.ca/about-apercu/business-plan-affaires_2012-2015-eng.htm. (accessed February 12, 2013).

Studman, B. (2010). A Collaborative Development Approach to Agile Statistical Processing Architecture – Australian Bureau of Statistics (ABS) Experience and Aspirations. UNECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS 2010). Daejeon, Republic of Korea, 26–29 April 2010.

Sundgren, B. (2007). Process Reengineering at Statistics Sweden. From UNECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS). Geneva, 8–10 May 2007. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2007/mtg1/wp.22.e.pdf. (accessed February 13, 2013).

Sundgren, B. (2010). The Systems Approach to Official Statistics. In Official Statistics Methodology and Applications in Honour of Daniel Thorburn. Stockholm University, Stockholm. 225–260. Available at: http://officialstatistics.files.wordpress.com/2010/05/bok18.pdf. (accessed February 12, 2013).

Sundgren, B. (2011). Towards a System of Official Statistics Based on a Coherent Combination of Data Sources, Including Surveys and Administrative Data. Available at: https://sites.google.com/site/bosundgren/my-life/Towardsasystemofofficialstatisticsbasedonacoherentcombinationofdatasources.doc?attredirects=0. (accessed February 12, 2013).

The Open Group (2009). The Open Group Architecture Framework (TOGAF), Version 9. The Open Group: San Francisco, CA. Available from: http://www.togaf.info/. (accessed February 12, 2013).

Todorov, V. (2011). Cooperation Models for Software Development. Meeting on the Management of Statistical Information Systems (MSIS 2011). Luxembourg, 23–25 May 2011.

United Nations Economic Commission for Europe, UNECE (2011). Strategic Vision of the High-level Group for Strategic Developments in Business Architecture in Statistics. ECE/CES/2011/1. Economic Commission for Europe, Conference for European Statisticians, Geneva, 14–16 June 2011.

United Nations Economic Commission for Europe, UNECE (2012a). Strategy to implement the vision of the High-level Group for Strategic Developments in Business Architecture in Statistics. ECE/CES/2012/10. Economic Commission for Europe, Conference for European Statisticians, Paris 6–8 June 2012.

United Nations Economic Commission for Europe, UNECE (2012b). The Generic Statistical Information Model (GSIM): Version 0.8. Available at: http://www1.unece.org/stat/platform/display/metis/GSIM+v0.8. (accessed February 12, 2013).

Vaccari, C. (2009). Common Reference Architecture (CORA) ESSnet: history and next activities, paper presented at the Meeting on the Management of Statistical Information Systems (MSIS 2009). Oslo, Norway, 18–20 May 2009. Available at: http://www.unece.org/stats/documents/2009.05.msis.htm. (accessed February 12, 2013).

Vale, S. (2009). The Generic Statistical Business Process Model, Version 4.0. Available at: http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model. (accessed February 12, 2013).

Wagner, J. (2008). Adaptive Survey Design to Reduce Nonresponse Bias. Unpublished Ph.D. dissertation, University of Michigan, Ann Arbor, MI.

Received May 2012
Revised November 2012
Accepted November 2012

Journal of Official Statistics, Vol. 29, No. 1, 2013, pp. 147–155

Discussion

Daniel Defays[1] *and Jean-Marc Museux*[1]

1. Preamble

This discussion draws mainly on four contributions in this special issue (those of Statistics New Zealand, NASS, RTI International and Statistics Netherlands) that report on experiences in integrating statistical production systems with a direct focus on industrialisation and integration-related issues. The framework proposed by Eltinge et al. for the integration of architecture and methodology is also taken on board. The article on the Real-Time Online Analytic System (Westat/NCHS) is marginally integrated into the discussion on data access.

2. Introduction: Systemic Issues

The articles from Statistics New Zealand, NASS, RTI and Statistics Netherlands all focus on efficiency gains through the integration of production systems. In most cases, the re-engineering of the systems that is necessary for integration leads to quality improvements through greater consistency and timeliness of statistical outputs. The systems also aim to give greater control to survey managers and the flexibility they need to optimise their processes. As a result, overall quality is better managed, opening the way for Adaptive Total Design as described by Eltinge et al.

Integration requires an overall framework (sometimes referred to as an architecture) for organising production systems. For methodologists and survey managers, architecture is a new concept which is worth close examination. The question "What is in an architecture?" is not easy to answer. What kind of new methodological issues does this raise? How is the concept of quality related to the concept of architecture, beyond the traditional quality of a survey or of a system designed to process information? Moreover, as the Statistics Netherlands article asks, how do we get from the concept of total survey design to total network design? Does this question our traditional definition of quality, which so far has been very much linked to a single data collection rather than to a set of data collections?

We will reflect on those issues through an analysis of the work carried out in different fora and described in the articles produced by Statistics New Zealand, NASS, RTI and Statistics Netherlands. Our approach is compatible with, and complements, Eltinge et al.,

[1] Eurostat, Directorate for Corporate Statistical and IT Services, Bâtiment Jean Monnet, L-2920 Luxembourg, LUXEMBURG Emails: Daniel.Defays@ec.europa.eu and Jean-Marc.Museux@ec.europa.eu
Acknowledgment: Authors wish to thank Walter Radermacher, Chief Statistician of the European Statistical System, for support and inspiring comments.

who proposes a framework for analysing systemic issues. We lay greater emphasis on a limited set of topics: the international perspective, the constraints imposed by architectures and standards, the problem of design/specification raised by modular approaches to data processing (called the "frame problem" in the rest of this discussion), some new attributes of architectures, and issues of implementation and data access.

3. The International Context

The integration of production processes belongs to a new set of initiatives aimed at industrialising the production of official statistics. This is acknowledged in Eltinge et al. among others, with a reference to international frameworks such as GSBPM (Generic Statistical Business Process Model) and GSIM (Generic Statistical Information Model). The business case for this kind of endeavour is clear and has been described several times: at the level of UNECE (United Nations Economic Commission for Europe) through the vision and strategic papers of a high-level group on business architecture for statistics (now renamed the highlevel group for modernisation of statistical production and services) (Statistics Netherlands 2011); at the EU level through a joint strategy developed to improve the efficiency of production and quality of European statistics (Eurostat 2010); and at national level, as illustrated in the articles under discussion as well as in countries such as Australia (Studman 2010), Canada (Doherty 2011), or Sweden (Axelson et al. 2011). Among the arguments put forward in favour of an integrated approach, there are the constraints put on resources which require greater production efficiency. This is recognised in all the articles in this issue as one of the main drivers of change. There are also users' expectations regarding the rapid delivery of information, or new data needs. Statistics New Zealand, for instance, has developed an environment where there is room for experimentation in testing scenarios and rapidly designing new indicators. Then there is globalisation, which requires the merging of different kinds of data from different origins, by improving the interoperability of national systems for example. As regards the proliferation of data sources, in some countries surveys are now used to complement administrative sources. There are also quality improvements brought about by the use of paradata to handle nonresponses better (RTI International) or to facilitate the use of mixed-mode data collection (NASS). A more complete list of drivers can be found in Radermacher and Hahn (2011).

4. The Concept of Architecture

Architecture is an abstract concept used in different contexts. All articles refer to it as an appropriate framework to deal with system complexity. This section explores some of the dimensions of this concept building on the articles contributions. The ISO/IEC defines an architecture as "the organisation of a system embodied in its components, their relationship to each other and to the environment and the principles governing its design". Eltinge et al. propose to qualify architecture through "the performance characteristics of the components, the integrated system, and system outputs". An architecture is a model for the design of business processes and related supporting information systems providing building blocks, principles and constraints aiming at facilitating the maintenance and the evolution of complex and open systems. An architecture encapsulates, *ab initio*, some knowledge and/or assumptions regarding the functioning and the organisation of the system. An architecture aims at striking the right balance between flexibility, rationalisation/efficiency, and quality. Information can be injected *a priori* in two different ways: either it is encoded in the models through parameters, input files, orchestration of building blocks or it is imbedded in the system itself. For instance, a building block can cover different GSBPM steps (say, A and B). In that case, the information that step A has to be followed by step B is encapsulated in the system/architecture and cannot be modified by the user. All architectures referred to in the articles build on the principle of modularity as the expression of the need for flexibility. The processing of information should be split into different steps corresponding to the different modules of the system. In the Statistics New Zealand article, statistical processing is carried out as an assembly of generic modules. NASS excludes multifunctional applications. Statistics Netherlands places modularity at the centre of its new system and RTI models a data collection as a combination of actions and steps, implementing those sequences in a standardised way to make them reusable. Eltinge et al. presents the increasingly used GSBPM framework as leading to highly modular systems. However, in doing so, the elementary blocks (representing operations, modules, GSBPM steps) out of which the processes are constituted represent the words of a language used to model statistical processing. Statistics Netherlands rightly questions the granularity of the modules, which, according to our analogy, means the vocabulary used to define processes. Eltinge et al. underline that in some cases standardisation may "prevent a given survey from making optimal use of survey-specific information" or "preclude some types of local optimisation of a methodological design". Regarding how the modules are defined, the articles do not go into sufficient detail, but it is clear that systems which lean on primitives such as exponential smoothing or linear interpolation or primitives such as historic imputation, seasonal adjustment (with a prescribed method encapsulated in the primitive), or micro aggregation (where the method for defining the micro aggregates from the original data is fixed) will have different "expressibilities". As defined for instance in Artificial Intelligence (see, for instance, Finlay and Dix 1996), an expressive system will be able to handle different types and levels of granularity of requirements. The more complex data processing it will enable to implement, the more expressive it will be. The range can span from a large set of statistical macros to a limited number of modules aligned to the GSBPM steps. A large number of modules gives expressibility and flexibility but makes the description of a process look like a computer programme. A system based on a limited set of modules will be more rigid and it will mean that a number of specific methods cannot be run. Nevertheless, it will probably lead to more standardised processes, which are easier to manage. There is no doubt that this is a key issue which should be carefully addressed with the right skills. In this context, it is important to draw on the competence of methodologists.

Similar issues arise when standards to represent the data and the metadata (including paradata) are embedded in the architecture. Here, too, the structure will restrict what can be represented. In SDMX (statistical data and metadata exchange), for instance, the notion of variable or question does not exist, contrary to what is offered by the DDI (Data Documentation Initiative). The articles under discussion do not go into much detail on how to represent key information. At European level, technical standards are currently being developed to structure, reference and process metadata (ESMS – Euro-SDMX Metadata Structure) and quality reports (ESQRS – ESS Standard for Quality Reports structure). Although like the modules they are extremely useful, they restrict what can be expressed. It would be interesting to explore if similar limitations exist in the way paradata are represented in Nirvana – the integrated data collection system designed by RTI.

5. The Frame Problem

Another difficult issue linked to the concept of processing as an assembly of modules is the implicit assumption that operations and data/metadata are independent of each other. More explicitly, a module is supposed to accept as input a set of information related to the data, metadata and parameters relevant for the processing to process the data and deliver an output which can be transformed into data, new metadata, or, for instance, a report on the processing. But as the processing is modular, in most cases the output in itself has no other *raison d'être* than to feed the next step of the process. But which step? If the modules are designed independently, and this is necessary in a modular approach, any kind of step which makes sense from a statistical or logical viewpoint is possible (imputation after validation, secondary cell suppression after the masking of primary confidential cells, etc.). However, how can we be certain that the output will be self-contained and ensure that all the relevant information for any kind of subsequent processing is available?

The traditional system from which GSBPM is derived provides an overview of the whole chain and the information needed at the different steps of the processing. In a modular and network approach, this is no longer the case. To take a simple example: when a value is imputed in an edit module, how should attention be drawn to it? By flagging the value as estimated? It may be that later on it will be important to know that this was estimated using a linear interpolation technique in order not to conclude to a linear trend which will be detected by an analysis module. You may well need to know even more about the method which has been used. Everything could be relevant *a priori*.

However, keeping the metadata/paradata of all information that might become relevant at some stage of a potential treatment is just not an option. This issue has similar features to the well-known "frame" problem in Artificial Intelligence. Here, the frame problem is the challenge of representing the effects of actions without having to represent explicitly a large number of intuitively obvious noneffects (see for instance Callan 2003).

6. The Quality of an Architecture

Methodologists are used to design surveys to optimise their overall quality under constraints (see for instance the interesting framework proposed by Eltinge et al.). This includes reducing the effects of nonresponse, frame coverage error, measurement and data processing errors while at the same time maximising timeliness and relevance. Often, other quality factors such as comparability and consistency are also taken into account (see for instance the Eurostat definition of quality). Strictly speaking, the last components are not characteristic of a survey, as they refer to the relationship which the statistics may have with statistics issued from other sources. They are, in fact, the property of a system, the kind of entity which becomes highly relevant when surveys are integrated. The shift in focus from a single survey to a set of integrated surveys affects the way quality has to be defined. In this new context, comprehensiveness (see the Dutch experience), redundancy, and local versus global optimisation become relevant issues. When the architecture is at stake, quality issues go beyond the quality of the system of integrated data collections it supports, as recognised in Eltinge et al.. Capacity is also relevant: to host new surveys (extensibility and adaptability); to promote reuse of components and interoperability (efficiency); to use standards; to be consistent with frameworks (such as accounting frameworks); to exploit the same set of data for multiple purposes; and to rationalise data processing in a sustainable way. This is an area which is only just starting to be explored by statisticians.

7. Process Standardisation Issues

In the model designed by Statistics New Zealand to meet the need to create or improve statistical output in a flexible way and make it a baseline activity for statistical subject-matter specialists, the common methods used to process the data (editing, outlier detection, imputation, adjustment of weights, estimation of model parameters, etc.) are combined into an end-to-end production process known as a configuration. In the Common Reference Environment (ESSnet CORE 2012), a proof-of-concept project carried out at European level to test the feasibility of modelling and orchestrating a statistical process as an assembly of services, work flow is also a key concept. When processes have to be integrated or reused, they need formal representations that can be processed by machines or by statisticians. Of course the notion of process is not new for official statistics and there are numerous methods for designing, managing and documenting them (see for instance the EPMS standard currently being drawn up at European level; Eurostat 2013). Nevertheless, the representation of processes as objects which drive processes as used in CORE and in Statistics New Zealand deserves to be highlighted. In the Dutch experience, the concept is slightly different, resting on the notion of a set of stable datasets (called steady states) and a production chain for the exchange of steady states. This goes beyond the notion of production lines and again only makes sense as part of a holistic approach.

More systemic approaches in statistical surveys and data collections will lead to new kinds of information concepts and objects that still have to be precisely defined and supported in production systems. These aspects were missing from GSBPM, but they have been partially integrated into the recently released Generic Statistical Information Model.

As stressed in Radermacher and Hahn (2011), "looking at individual production processes and grouping them into larger entities is not enough. All processes will need to be tackled in a holistic approach with a view to enabling their integration and modularity within the whole statistical authority, across statistical authorities and with international organisations. Thus, there is a need to enlarge the scope of standardisation with meta-processes – going beyond the GSBPM – that allow for a bird's eye view on production". Figure 1 illustrates the scope of a statistical "factory":

In this broader context, an architecture can be extended to cover nonproduction processes such as common professional standards for statistics as developed by the European Statistical System project of an European master programme of official statistics, standardised press releases and other press activities, and so on.

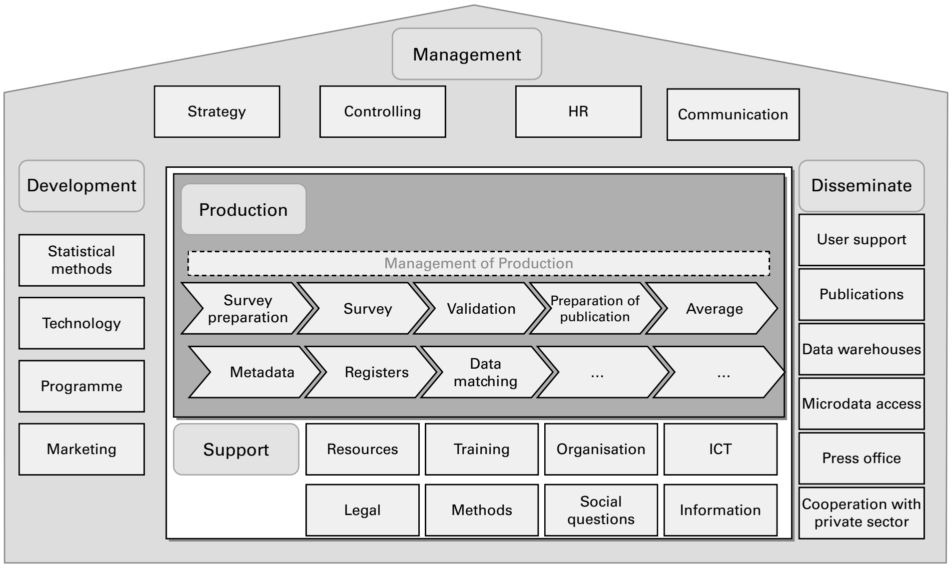

Fig. 1. The statistical factory (Radermacher and Hahn 2011)

Finally, the standardisation of processes/metaprocesses cannot be disentangled from a "product"- component, which is the second strand developed in Radermacher and Hahn (2011). As long as we define our products in a tailor-made way, fit for (and only for) the purpose of specific users, the industrialisation-potential of processes will be extremely limited. The new architectures for Official Statistics should include the building blocks for "multipurpose-products", which are ideally "*prêt-à-porter*", meaning of high quality, usable for many purposes, and provided in the form of a coherent set of mesodata organised through accounting frameworks.

8. The Implementation Issue

Although the industrialisation of statistical processes and the rationalisation of information systems have been discussed for several years, success stories and implementation reports are still rare. On the basis of the four articles, good practices can be derived for the re-engineering production systems.

All the initiatives target a paradigm shift in the production environment. This requires high-level skills, dedicated governance mechanisms, an adequate mobilisation of resources and deep involvement on the part of the organisation business process owners.

All transformation initiatives include well-defined governance that breaks down hierarchical barriers. Architecture boards are frequently required for translating strategic goals into models, operationalising the goals and balancing local and corporate interests.

The ongoing experiences described in the articles show that while the final strategic goal is very important, it is nonetheless essential for the programme to be structured as a sequence of transitions towards intermediate states, where benefit can be demonstrated and reorientation is possible should there be changes in the environment. Success stories are frequently built on past failures which have been analysed and served to build capabilities.

Resource allocation must be agile as it can be necessary to reallocate resources (be it financial or human) rapidly at any stage of the implementation. This is difficult to integrate in a standard cost-benefit analysis. A recommended way (Statistics New Zealand) is to proceed step by step (phasing), measuring the benefit of specific pilots, either further optimising the future state or continuing the planned implementation.

The articles highlight the need for the strong involvement of users. Agile development involving users, developers, and designers is also proposed as a means for the gradual creation of competence.

The skills for designing a new system are not always available in house. In the event of outsourcing, shadowed capabilities (RTI International) have been found to be necessary in order to build in-house competence to ensure the sustainability of the system.

Eurostat has recently designed a programme for developing a set of common infrastructures to support the implementation of the ESS vision for the production of EU statistics (Eurostat 2009). The main technical infrastructure components include: 1) a secured network supporting the flow of information among ESS partners; 2) a platform for sharing common services; 3) common data repositories; and 4) a common validation architecture. The approach combines the business realisation of key flagship initiatives which will demonstrate and pilot key features of the ESS Vision (exchange of information and interoperability of statistical processes) while gradually building up the infrastructure components of the future system. The approach involves a paradigm shift from a stovepipe, bottom-up and project-based approach to a programme-based approach closely tying together business outcomes and crosscutting infrastructure developments.

The programme infrastructure being set up builds on the above principles adapted to the international and decentralised context. For more information on current developments, see Eurostat (2012).

9. The Data Access Issue

In a systemic context, as acknowledged in the business case for official statistics in a global information market (Radermacher and Hahn 2011), access to detailed and confidential information by partners or external users has to be integrated into the architecture design. Official statistics have traditionally maintained a privileged relationship with information providers. Indeed, the protection of individual information collected for statistical purposes is usually seen as the foundation for the quality of the primary information collected. This has led to a well-developed security and statistical disclosure limitation architecture components. However, it is also building barriers against the wider integration of statistical systems and the necessary specialisation of statistical work. The move towards reusing existing information (currently coming mainly from administrative sources, as shown by the Statistics New Zealand article) is creating a slight shift in the focus of statistical confidentiality, to the extent that there will be less direct interaction between statisticians and data providers, although the principle of respect for confidentiality will have to remain enshrined in the future architecture. Given the place of data analytics in the new architecture, innovative solutions must be found to ensure the appropriate trade-off between flexible access to detailed data and statistical disclosure limitation. The Westat/NCHS article gives a detailed account of the design of an online analytic system targeting external users, which is required in order to cope with statistical confidentiality. In an integrated system such as the European statistical system, new architecture should push the provision for security towards the system's border, implementing the Schengen model for the movement of individuals within the EU (see for instance Bujnowska and Museux (2011) on the implementation of the Schengen model for research into microdata access). Under the Schengen approach, all the data collected under European legislation should be considered as the common good of the European Statistical System and, consequently, each national authority should be authorised to access and grant access to all European data. This could only be possible if a common architecture defines rights and duties, protocols and basic principles (most likely to be enshrined in the appropriate legislation).

In the medium to long term, the public's experience of statistical confidentiality is expected to change, given the ever-increasing amount of individual information deliberately or unconsciously made available on the internet through social networks or mobile devices. Statistical confidentiality will no doubt have to be redefined to take a changing reality into account. A new paradigm will almost certainly have to be found.

Efficiency gains, quality improvements, and higher reactivity push statisticians to conceive their systems in a more holistic way. This requires new concepts; this raises new issues. The shift from systems to architecture, from local optimisation to global optimisation, from the production chain to the factory is on its way. The current empirical approach needs to be embedded in a more theoretical framework. This is a new challenge for statisticians.

10. References

Axelson, M., Engdahl, J., Fossan, Y., Holm, E., Jansson, I., Lorenc, B., and Lundell, L.G. (2011). Enterprise Architecture Work at Statistics Sweden. Paper presented at the 2011 International Methodology Symposium, Statistics Canada.

Bujnowska, A. and Museux, J.-M. (2011). The Future of Access to EU Confidential Data for Scientific Purpose, 2011 UNECE-ESTAT Work Session on Statistical Data Confidentiality.

Callan, R. (2003). Artificial Intelligence. Basingstoke: Palgrave Macmillan.

Doherty, K. (2011). Update on How Business Architecture Renewal is Changing IT at Statistics Canada. Working paper 3, MSIS 2011, United Nations Economic Commission for Europe.

ESSnet CORE (2012). Common Reference Environment, EU Commission funded project. Available at http://www.cros-portal.eu/content/core-0 (accessed February 14, 2013).

Eurostat (2009). Communication 404/2009 of the Commission and the European Parliament on the "Production of EU statistics: a vision for the next decade". Available at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0404:FIN:EN:PDF.

Eurostat (2010). Communication from the Commission on the European Parliament and the Council on the Production Method for EU Statistics — Joint Strategy Paper, 5th Meeting of European Statistical System Committee, document ESSC 2010/05/6/EN.

Eurostat (2012). The ESS VIP Programme, 15th Meeting of European Statistical System Committee, document ESSC 2012/15/6/EN.

Eurostat (2013). Better Documenting Statistical Business Processes: the Euro Process Metadata Structure, forthcoming paper METIS UNECE Work Session, May 2013.

Finlay, J. and Dix, A. (1996). An Introduction to Artificial Intelligence. London: UC Press.

Radermacher, W. and Hahn, M. (2011). The Business Case for 21st Century Official Statistics. Unpublished paper, available at http://www.cros-portal.eu/content/business-case-21st-century-official-statistics (accessed February 19, 2013).

Statistics Netherlands (2011). Strategic Vision of the High-Level Group for Strategic Developments in Business Architecture in Statistics. Paper presented at the 59th plenary session of the conference of European statisticians, UNECE.

Studman, B. (2010). A Collaborative Development Approach to Agile Statistical Processing Architecture — Australian Bureau of Statistics (ABS) Experience and Aspirations. Working paper 3, MSIS 2010, UNECE.

Journal of Official Statistics, Vol. 29, No. 1, 2013, pp. 157–163

Discussion

Alan F. Karr[1]

These five articles and the framework proposed by Eltinge, Biemer, and Holmberg illuminate an exciting view of the future of official statistics. The articles are innovative, forward-looking, and remarkably complementary to one another. Below, for brevity, I refer to the five articles as the NASS, StatsNL, StatsNZ, RTI and Westat/NCHS articles. I will comment on each article individually, on them collectively in the context of the framework, and on the entire concept of "systems and architectures for high-quality statistics production." First, however, I present two complementary perspectives on official statistics.

1. A Broad Perspective

Official statistics are like cars. A few aficionados are interested in them for the sake of style, technology or snob appeal, but almost everyone else sees them solely as a means of fulfilling other needs. Utility and affordability affect people's choices of cars, and in the future, if not already, will affect people's choices of official statistics. These five articles and the Eltinge/Biemer/Holmberg framework are nascent recognitions of that reality.

Except for respondents, many people associated with the production of official statistics – myself included – are aficionados. We think of data as the end product – and the more stylish, technologically advanced and elegant, the better. Most other people, if they were to think about it, would say that, like a car, data are valuable (if at all) because they enable other needs to be met. In reality, the end product of our efforts is the decisions made by governmental bodies on the basis of official statistics, at national, state/provincial and local levels, as well as decisions by private sector organizations. Examples of the latter are decisions where to locate manufacturing and retail facilities.

Some people might also include as an end product improvements to society resulting from (academic and other) research that employs official statistics. Education, health and transportation are examples of contexts where such research has clearly made a difference, although widespread recognition that this is so may be lacking. Note carefully that I did not term the research itself a "true end product," but rather the societal impact of the research, which places data two steps removed from reality.

Only boutique car manufacturers are run by aficionados, and official statistics agencies run by aficionados risk becoming enamored of the product, rather than of what it enables. Ultimately, our value to society depends on the quality of decisions and the quality of societal improvements. These are not the same thing as data quality, about which we still know precious little, so the challenges are immense.

[1] National Institute of Statistical Science 19 T.W. Alexander Drive, P.O. Box 14006 Research Triangle Park, NC 27709-4006, U.S.A. Email: karr@niss.org

Confronting the challenges requires new modes of thought, organization and operation. Maintaining political and scientific impartiality but being conscious of the political and sometimes partisan decisions based on the data we produce is not impossible – only exquisitely difficult. Failing to understand and serve our customers is a death sentence. Modernizing our systems and architectures, in ways described in these articles and other ways, is an essential step away from the precipice. The authors and editors are to be congratulated for recognizing this and doing something about it.

2. A Narrower Perspective

Speaking as an aficionado, like most others in official statistics, I would include in my list of the biggest obstacles (but also opportunities!) that we face: (1) Declining agency budgets, (2) decreasing response rates, and (3) the need to use administrative data as a complement or substitute for survey data. Importantly, the seeming gulf between these and the broader perspective in Section 1 is not unbridgeable. Indeed, the systems and architectures described in these five articles represent thought and action that begin to unite the two perspectives.

Perhaps not everyone will agree with this analogy, but system- and architecture-based production of automobiles serves both customers and producers. It improves the product, makes it more broadly affordable, and at the same time focuses intellectual creativity on the issues that really matter: price, features, reliability and safety. Perhaps there has become less room for artisans, but this is more than compensated for by the new opportunities. These five articles embody the same kind of thinking: production of official statistics can become less artisanal without becoming less professional, less exciting or less rewarding. On the contrary, new systems and architectures may be the best if not last chance for us to address the issues that really do matter, and that really do demand that we be professionals.

3. The Five Articles

The NASS, StatsNL and StatsNZ articles describe agency-wide efforts that share some characteristics, but also differ in important ways. The RTI article focuses exclusively, albeit broadly, on survey operations, and the Westat/NCHS article addresses the still more specific issue of data confidentiality within a systems and architectures context.

NASS. This article, which in my opinion is outstanding, describes a near-complete makeover of an official statistics agency for which new IT infrastructure was the enabler, but not the purpose. The effort was initiated by the CEO of the agency, who clearly built up significant support within NASS for it, but who also seemingly made clear that changes were necessary. Three factors appear to have driven the effort: reduction of personnel costs in response to declining budgets, modernization of – especially the software component of – the IT infrastructure at NASS, and reorganization of a legacy system of decentralized field offices into a more efficient, less cumbersome structure.

Although the NASS article emphasizes survey operations, the scope of the changes is clearly broader, especially with respect to internal and external dissemination of data and analyses. The article demonstrates the thought, planning, and care that went into the creation of the new architecture, including the need for a business case for each component of it.

Among the most interesting aspects of the process is the way in which NASS combined in-house expertise with that of contractors. NASS depends less on contractors than some other U.S. official statistics agencies for collection of data, but clearly recognized and responded to the need for external assistance. Also outstanding is the people-oriented manner in which the process was carried out, as confirmed by the inclusion of references such as Collins' *Good to Great* (Collins 2001), which I have read and recommend highly. (Another of my favorite references of this sort will be mentioned in Section 5.) Not surprisingly, most of the lessons learned presented in this article have to do with people, not systems and architectures. They were possibly more foreseeable than the article seems to suggest, but at least they have been learned.

What to me is missing from this article is the "end products of our efforts" perspective in Section 1. It is not clear to what extent users of NASS data were involved in the process, nor what role they are playing in evaluation of the changes. Are better decisions being made as a result of the new architecture? Is better research being conducted? How will NASS know?

StatsNL. The redesign of Statistics Netherlands' production system described in this article is comparable in scope to that of NASS, but with more emphasis on the integration of data from registries. There are many similarities between the two efforts. Like NASS, StatsNL began from strategic rather than operational goals. Like NASS, it created advisory and governance structures to oversee the process. And like NASS, it was responding to cost considerations, although in this case the main concerns seem to be costs of system implementation and maintenance. And one final similarity: both StatsNL and NASS sought explicitly to reduce respondent burden, albeit in different ways.

However, there are also important differences. StatsNL created and based its system architecture on a rather formal data model, as well as a set of use cases. StatsNL may have realized earlier in its process that, as the article states, "the architecture itself does not make methodological choices." The StatsNL and NASS perspectives on edit and imputation appear to differ, although there is not sufficient detail in the articles to say precisely how. The StatsNL view of costs and benefits appears to be somewhat broader than the NASS view, but this may reflect in part the relative maturities of the two projects. StatsNL does seem to have thought more directly about risks, as well as to have realized that no matter how detailed the planning, sound the execution and thorough the evaluation, things will not be right the first time.

The extent to which StatsNL relied on external expertise during the redesign process is not clear. Whether its system is robust enough to withstand the future remains to be seen.

StatsNZ. The context, systems and approach described in this article are generally more similar to those for StatNL than those for NASS, but there remains much in common among all three efforts. StatsNZ is very much like NASS with respect to breadth: "Stats 2020 [. . .] involves changes in the way the organization works; who it works with and how; to the systems, tools, and processes used; and to the skills, and behaviours needed for success." Like StatsNL, StatsNZ is highly focused on use of administrative data, a luxury not equally available to agencies in other countries.

Among the strengths of this effort are explicit consideration of risk (e.g., arising from legacy software) and the user-centric "quality framework and quality indicators." However, the view of quality remains that of an aficionado. The nine principles laid out in

Section 2.2 of this article are laudable, but at the same time are to me too general to be actionable and too dogmatic to be useful. In the same way that there is no disclosure prevention but only disclosure limitation or avoidance, saying that "Surveys will be used only to fill the gaps that cannot be met from administrative sources" misses the point, especially in countries like my own where distrust in the national government is high.

This article raises a number of technical issues that are interesting and important. Versioning of data is, I believe, largely ignored by official statistics agencies. No serious producer of software could exist without version control. In a project with a major telecommunications company to perform statistical analyses of their software development processes, I worked with a version control system that enabled the software to be "built" as of any date in the twenty-year history of the system. Should we not think in the same way about data? Steps in this direction are occurring: the U.S. National Center for Education Statistics (NCES) will in the not-distant future release a thoroughly modernized dataset for its 1982–92 High School and Beyond (HS&B) study (NCES 2013) that nevertheless permits replication of every previously published analysis.

StatsNZ identifies clearly the need for automated data editing. StatsNL and NASS recognize the same need, but even though there is progress, no one has yet developed the mature, powerful tools that we really need – especially to integrate editing, imputation and statistical disclosure limitation.

StatsNZ's configuration store is also important. In the software development project alluded to above, more than one-half of the 200 million+ lines of code were in make and header files. As statisticians, we should be comfortable with the concept of a configuration store, which is essentially metadata for software.

RTI. RTI is a survey contractor, conducting surveys on behalf of official statistics agencies in the U.S. While efficient in some ways (RTI, NORC, Westat and other private sector organizations have incredible accumulated knowledge, and are, as evidenced by this article, true innovators), this arrangement creates two disconnects. First, the agency loses direct contact with respondents, contact that NASS, for instance, finds invaluable. Second, the survey contractor has no (or little) direct relationship with data users, which can lead to data quality being measured too mechanically (e.g., by response rates or contact attempts necessary), rather than in the user-centric perspective in Section 1, or by metrics associated with inferential uses of the data, let alone in terms of decisions based on the data.

The Nirvana system described in the RTI article is a tool for survey managers concerned on a day-by-day basis with issues such as response rates, interviewer assignment and costs. It is overtly a tool to support operational decisions. In this respect, for instance, through the now beaten-to-death dashboard metaphor, it is similar to some components of the NASS system. More important, however, is the data and metadata infrastructure on which it rests. The standardization, modularity and extensibility are entirely in the spirit of NASS, StatsNL and StatsNZ. Whether as a result of proprietary or other concerns, the level of detail about the system is less than in the other articles.

The account of the development process, on the other hand, is as interesting as it is instructive. Like NASS, StatsNL and StatsNZ, RTI understood from the beginning the people-centricity of the system development process, and put in place what seem to have been very effective advisory and decision-making mechanisms. However, three groups seem absent: RTI's clients (for whom it conducts surveys), respondents (whom NASS, StatsNL and StatsNZ have involved in their system development projects) and data users. Why?

Westat/NCHS. Westat, like RTI, is a survey contractor. This article, unlike the RTI article, describes a partnership between Westat and a U.S. official statistics agency – the National Center for Health Statistics (NCHS), which is part of the Centers for Disease Prevention and Control (CDC). It is not an article about systems development *per se*: the dissemination system that motivates it does not exist and might never exist.

The article is, instead, an object lesson in the external realities of system building: enterprise architectures, standardization, modularity, extensibility, virtual servers and APIs aside, how do we determine system requirements, and how do we know that they can be, or when they are, fulfilled?

No amount of generalities can answer these questions, and so this article stands as a beautifully concrete instantiation of the problem. Without going into technical detail, the basic problem is whether a query-based access system (Karr et al. 2010) can run in real time on both restricted (confidential) data and publicly released versions of the same data. None of the other articles addresses a usability issue of this specificity; indeed most of them consider dissemination only obliquely, and confidentiality not at all.

As in the other articles, there is a gap between system and users, let alone decisions, research, or societal improvements. There is little discussion of data utility even from an inferential perspective, although a lot is known about utility for tabular data (Cox et al. 2011; Gomatam et al. 2005). The discussion of risk omits consideration of nefarious users to whom denial of a query may be informative, or who have the computational resources and wit to enumerate all tables compatible with the released information.

The extremely important message of this article is that implementation of systems and architectures has a detailed as well as high-level side, which can be extremely delicate. The devil is always in the details. Bringing them in too soon can stop a project in its tracks. Bringing them in too late, on the other hand, can be fatal. Most software projects that fail do so because of incompletely, inconsistently, or incorrectly specified requirements. The NASS, StatsNL, StatsNZ, and RTI articles are all, to varying degrees, vague about how requirements were specified and modified. The authors of this article deserve praise for reminding us how much requirements matter.

4. The Five Articles in the Context of the Eltinge/Biemer/Holmberg Framework

In a related setting (Cox et al. 2011) I and co-authors have criticized risk-utility paradigms in statistical disclosure limitation as useful for "how to think, but not how to act." By this, we distinguish frameworks for structuring decisions from systems and architectures for making decisions. To a significant degree, I feel the same about the editors' framework. Even though they would be first to admit that the framework is not in remotely final form, at the current level of generality, there is little clarity how to move forward. Albert Bowker, former chancellor at the University of Maryland and the University of California Berkeley, once told me that it is often more important to take a step in the right direction than to be able say what the final destination is. Eltinge, Biemer, and Holmberg might do well to heed his advice. The authors of the five other articles certainly have: they all recognize the difference between a step and a destination.

The most glaring omission in the framework is people. Each of the five articles emphasizes the centrality of people to systems and architectures. To quote from StatsNZ, "Systems can complement, but are not a substitute for, [this] intellectual and human productivity." The equations in the framework do not even begin to capture the complexities of the people immediately involved in the process, let alone data subjects, clients or users, researchers, decision makers, or the public at large.

To write down equations of the form $Q = g_Q(X, Z, \beta_Q)$ simply does not move the conversation forward. Why should there be a parameter β_Q when g_Q is entirely unspecified? What if every object involved is of unimaginably high dimension? What if the "science" is nothing but a collection of special cases? At this stage, to suggest particular functional forms or methods of statistical estimation seems premature, and deflects attention from the central issue of better decisions.

5. Conclusion

The stunning common characteristic of the systems and architecture projects reported here is that they are all, perhaps to degrees not yet known fully, successes. To the agencies and people involved, I can only say "Bravo!" In an age of highly publicized, expensive software projects in both the public and private sectors that are failures to the point of abandonment, it is truly heartening to see what is possible with limited but intelligently applied financial, human, hardware and software resources. Repeating myself, these five articles demonstrate how much can be accomplished given attention to people and the willingness to take a step in the right direction.

So, do these successes mean that the situation for official statistics is now rosy? Sadly, I fear that it is not. These five articles are, at the moment, rather isolated instances of success. There have also been failures, many of whose lessons learned were never shared with others. The corporate-like thinking associated with some systems and architecture projects may not go far enough. In addition to *Good to Great*, my favorite business reference is Clayton Christensen's *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail* (Christensen 1997). Christensen originated the concept of *disruptive technologies*, such as personal computers, that vitiate old-line companies "from below," with products that are both better and cheaper, or are so much cheaper that being as good as (or even almost as good as) the incumbent technology leads to success.

The business and profession of official statistics are vulnerable to disruptive technologies. What if data from Facebook, Google, and Twitter are good enough to support the decisions that need to be made, as well as the research that is necessary to improve, say, public education and health? What if privatization of government data collection were to work "well enough," and at one-tenth the current cost?

Haughty sneers that we more fully respect confidentiality or produce higher quality data may not help. (American Motors once thought that it made better cars than General Motors, Ford or Honda, and RIM dismissed iPhones as frivolous alternatives to Blackberries.) These five articles, and the Eltinge/Biemer/Holmberg framework, show that disruptive technologies can be harnessed, and should inspire others. However, they create no conceivable case for complacency.

6. References

Christensen, C.M. (1997). The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail. Cambridge, MA: Harvard Business Review Press.

Collins, J. (2001). Good to Great: Why Some Companies Make the Leap . . . and Others Don't. New York: HarperBusiness.

Cox, L.H., Karr, A.F., and Kinney, S.K. (2011). Risk-Utility Paradigms for Statistical Disclosure Limitation: How to Think, but not How to Act (with discussion). International Statistical Review, 79, 160–199.

Gomatam, S., Karr, A.F., and Sanil, A.P. (2005). Data Swapping as a Decision Problem. Journal of Official Statistics, 21, 635–656.

Karr, A.F., Kinney, S.K., and Gonzalez, J.F. Jr. (2010). Data Confidentiality—the Next Five Years: Summary and Guide to Papers. Journal of Privacy and Confidentiality, 1, 125–134.

National Center for Education Statistics (NCES) (2013). High School and Beyond (HS&B). Information available at: http://nces.ed.gov/surveys/hsb/.

Journal of Official Statistics, Vol. 29, No. 1, 2013, pp. 165–169

Discussion

Frauke Kreuter[1]

This special issue on "Systems and Architectures for High-Quality Statistics Production" is a stimulating resource for statistical agencies and private sector data collectors in a challenging time characterized by massive amounts of data, from a variety of sources, available in varying intervals, and with varying quality.

Traditionally, statistical products were created from a single source, most often through surveys or administrative data. However, neither surveys nor administrative data alone can match the data needs of today's society. In addition, the need to reduce the costs of data production necessitates that multiple sources are used in combination. The need to reduce costs also necessitates the streamlining of production cycles, and the increasing difficulties in data collection itself require such systems to be much more flexible than they have been in the past. Increasingly, these reasons are driving statistical agencies and private data collectors to redesign their entire data production cycle. The examples in this special issue from Statistics Netherlands and Statistics New Zealand demonstrate such developments in government agencies; the example from RTI reflects efforts visible among private sector data collectors. This commentary will highlight some issues of general interest related to organizational challenges, and some that create the basis for reproducible research and are therefore of general interest to the research community.

1. Organizational Challenges

A common phenomenon among many national statistical agencies and also many private data collectors is an *organizational structure around projects or topics*, rather than around tasks. Such an organizational structure is often referred to as one involving stovepipes or silos, and all three articles comment on the presence of such structures in their organizations. In each of these organizations, the *silo structure* resulted in a multiplicity of software solutions, duplication of work, a lack of dissemination of knowledge from silo to silo on particular tasks, and a lack of standardization in survey products and documentation.

To overcome the disadvantages of silos, each organization restructured the statistics production process. At RTI, the restructuring revolved around a core of software tools that in the future will be used across all survey data collection projects as well as a common data storage system. Similar approaches were taken by Statistics New Zealand. Statistics Netherlands went a step further and also changed the structure of the organization itself,

[1] Joint Programme in Survey Methodology, University of Maryland, Ludwig Maximilian University, Munich and Institute for Employment Research (IAB), Email: fkreuter@survey.umd.edu

implementing modular production lines, and moving away from parallel production lines. For Statistics Netherlands, the modular approach allowed for better use of scarce resources, such as expertise in specialized fields, like sampling. Moving away from silos also greatly facilitated the use of the same metadata (data about data) and documentation across systems, streamlining training needs and ultimately reducing costs.

All three organizations report on the difficulties they encountered in implementing the changes to existing data production systems. Most of these challenges are not unique in this context and can be found elsewhere. One of these challenges is the *adoption of a new system during ongoing production*. More subtle, but nevertheless important problems reported in these articles are those arising from the *lack of participatory decision making*. Because individual units are most familiar with specific processes, input from these units is essential to developing new routines. The Statistics Netherlands article points out that communication across people working in Information Technology and survey production is crucial to a successful operation. Another common challenge is the continuous updating of the data within the systems and the availability of data from the system to each production unit at appropriate times.

All three articles, but in particular the two from the national statistical agencies, emphasize the growing demand for *the use of administrative data* on its own or in combination with survey data. However, the descriptions in the articles could spend more time detailing on how such combinations are achieved. When data from multiple sources are combined, individual records do not always have unique and identical sets of identifiers. In such cases, probabilistic record linkage is used. Privacy concerns often result in identifiers being encrypted, creating another challenge for unified systems. Here, new methods are being developed to allow for privacy-preserving record linkage (see e.g., Schnell et al. 2009). In some cases, clearing houses (within or across agencies) are used to allow for record linkage with unencrypted identifiers. It would be interesting to know how the various agencies handle this problem.

When administrative data is used for statistic or data production, its *quality* is of paramount importance. The three articles here lack some clarity on how the quality of administrative data is assessed and ensured beyond edit checks. It is not unlikely that several data sources will have overlapping sets of variables and that those variables are of different quality, depending on their source and the importance of a given variable when the data are entered. Incorporating the perspective of efficient data linkage into the building of statistics production systems is important.

2. Reusability and Reproducibility

One of the key motivations for the restructuring of the statistical production process is the possibility to reuse existing knowledge, data and code across statistical products and across organizational silos. The focus of all three organizations on employing unified software systems and comparable documentation systems not only enhances reproducibility but creates the basis for quality assurance. For example, at Statistics New Zealand, undocumented manual repairs to the data threatened the quality of the final product in the past. The article comments on new software systems that allow automatization of routine tasks (to avoid hand-coding) and the automated creation of data tables for storage and reports. Such automatization is often missing, and even organizations with smaller projects would benefit from unified software code stored for long-term use and reproducibility.

Metadata play an important role in this process. In order to move data across silos or production processes in an automated fashion, certain standardizations must be in place. One example of this kind of standardization is the introduction of common labels for corresponding variables. As simple as this might seem from the outside, each article commented on the difficulty of achieving such standardization. With long-standing data products in particular, such changes can be tedious for individual silos. Having achieved a common set of metadata that allows easy exchange of data within the organizations is therefore a major accomplishment.

However, this agenda could be pushed even further. Thinking of statistical data users located outside the statistical agencies, standardization across statistical offices would also be advantageous. Outside users are much more likely to use data from various sources, maybe even various countries, than users within the same agency. Outside users often attempt to combine or compare such data. Large resources are then needed (over and over again) when researchers make data comparable *post hoc*. The standards developed by the Data Documentation Initiative (DDI 2012) could help overcome this problem. DDI has developed standards for describing data from the social, behavioral, and economic sciences that support the entire research data life cycle, including the collection, processing, distribution, and archiving of data. For example, a set of controlled vocabularies sets metadata standards in semantics (which elements should be defined), and content (what values should be assigned to elements and how). A naïve user of data might think, for example, that all surveys ask questions about race and ethnicity in the same way, whereas quite the opposite is true even within a single organization. To date, primarily academic entities have subscribed to the DDI standards, but some Research Data Centers (RDC) affiliated with government organizations, such as the RDC at Statistics Canada or the RDC affiliated with the German Institute for Employment Research (IAB), are in the process of designing their metadata systems according to such international standards (Bender et al. 2011).

Of course, if analysts frequently shared their data preparation and analysis code, efficiency gains in such repetitive work could be accomplished even without the international standardization of metadata. Fortunately, some journals already encourage, if not require, the dissemination of analysis code. However, the inability to search journals' supporting online material make the dissemination of analysis code through free code repositories such as GitHub (http://github.com) and SourceForge (http://sourceforge.net) much more desirable (for a larger discussion, see Peng 2011). Publicizing the code used by statistical agencies to transform raw data into data products can not only increase efficiency, but also contributes to quality assurance by making the process reproducible.

In addition to metadata and shared analysis code, paradata – data created as a by-product of the production process – are another key element in quality assurance, and one that has to be considered in the construction of unified systems. Many agencies are interested in using survey production-based paradata to inform decisions about cases during ongoing fieldwork and to allow for a timely shift of cases from one production mode to another. One example – beyond the ones discussed here – is the U.S. Census Bureau, which is currently creating a flexible and integrated system to use paradata in ongoing survey management (Thieme and Miller 2012). However, paradata can also be used to monitor the processes within the agencies themselves. Paradata examples of the latter could be records on edits done to a particular case or data set (already prominently used at RTI), production times for a particular module (e.g., in the Dutch case), or the number of data manipulations necessary to move raw data into long-term storage in flexible data bases (e.g., in the case of New Zealand). In this sense, the whole statistical production process can be monitored through paradata. The development of useful paradata is currently still in its infancy and their measurement error properties are unknown (see also Kreuter 2013).

3. Outlook

It is interesting that each organization developed its own system, and did not seem to collaborate with others in the design process. This special issue is a valuable step towards fostering such collaborations. These articles will help disseminate information about the new system developments and hopefully spur dialogue and interaction amongst the various data collection entities. My wish would be for organizations to share some of the system infrastructure when possible. While it would probably be difficult to share the entire IT infrastructure, code pieces could very well be provided and would be of use to others – for example, code pieces necessary to extract paradata from data collection software, or algorithms used in call scheduling or models used for responsive designs. A larger discussion would also benefit from developments in other fields or those that develop across disciplines. To mention just one example, the British company Jisc aims to create a Virtual Research Environment (http:jisc.ac.uk/whatwedo/programmes/vre.aspx).

The discussion about the system development also highlights anew the need for people trained with a total survey error perspective. The total survey error perspective (TSEP) deconstructs error in a survey estimate into its constituent pieces, such as nonresponse error and errors due to measurement. Being aware of the various sources of error and their likely contributions to a given estimate based on previous research will allow for informed trade-off decisions as data are combined within the system. A framework similar to the TSEP could also be used to understand the errors in administrative data, although much more work is needed in this area. One obvious place to look to begin to create an error taxonomy for administrative data is the study of questionnaire design, for administrative data are often also collected through self-administered forms that resemble surveys.

Survey methodologists can also help identify the effects of changes made to one production module on other modules. This will be important, for example, when unified systems are used to move cases in a flexible way, mid-field, from one mode of data collection to another. In addition, survey methodologists can use their tools to assess the possibilities of merging various data sources with different error structures (a point not mentioned in the three articles), and their expertise in the analysis of such linked data (Lahiri and Larsen 2005). Finally, survey methodologists can contribute their knowledge about issues of data linkage that need to be addressed at the data collection stage, for example, when asking for linkage requests (see e.g., Calderwood and Lessof 2009; Sakshaug et al. 2013).

4. References

Bender, S., Dieterich, I., Hartmann, B., and Singula, D. (2011). FDZ-Jahresbericht 2009/2010. Available at: http://doku.iab.de/fdz/reporte/2011/MR_06-11.pdf (accessed February 14, 2013).

Calderwood, L. and Lessof, C. (2009). Enhancing Longitudinal Surveys by Linking to Administrative Data. Methodology of Longitudinal Surveys, P. Lynn (ed.). New York: Wiley.

DDI (2012). What is DDI? Available at: http://www.ddialliance.org/what (accessed February 14, 2013).

Kreuter, F. (ed.) (2013). Improving Surveys with Paradata: Analytic Use of Process Information. New York: Wiley.

Lahiri, P. and Larsen, M. (2005). Regression Analysis with Linked Data. Journal of the American Statistical Association, 100, 222–230.

Peng, R. (2011). Reproducible Research in Computational Science. Science, 334, 1226.

Schnell, R., Bachteler, T., and Reiher, J. (2009). Privacy-Preserving Record Linkage Using Bloom Filters. BMC Medical Informatics and Decision Making, 9(41).

Sakshaug, J., Tutz, V., and Kreuter, F. (2013). Placement, Wording, and Interviewers: Identifying Correlates of Consent to Link Survey and Administrative Data. Survey Research Methods (forthcoming).

Thieme, M. and Miller, P. (2012). The Center for Adaptive Design. Presentation to the National Advisory Committee on Racial, Ethnic, and Other Populations. October 25, 2012, Washington, D.C.

Journal of Official Statistics, Vol. 29, No. 1, 2013, pp. 171–175

Discussion

Thomas Lumley[1]

1. What is the Same? What Has Changed?

Collecting and integrating high-quality information from a census or population sample will always be difficult and expensive, requiring specialized expertise, and benefiting from accumulated institutional memory. The same used to be true of analyzing the data, which required expensive computers with many megabytes of storage and advanced software capable of computing appropriate standard errors for the sampling design.

In the modern world, however, any major general-purpose statistical package will support not just tables but regression models on complex samples, and ordinary laptops have the processing power to handle almost all complex survey data sets. For the few very large samples, such as the American Community Survey (3 million/year) and the US Nationwide Emergency Department Sample (25 million/year), it is still possible to perform analyses with standard statistical software on commodity server hardware: for example, a rack-mount Linux server with 128 Gb memory can easily be found for under US$5,000.

The statistical expertise needed for analysis has also become less specialized. The software only requires users to be able to understand and describe the basic design, or a one-stage approximation to it, and then to use essentially the same scripting syntax and analysis options as they do for unstructured data. This is perhaps most dramatic in Stata, where the svy: prefix can be applied to almost any command, but is also true of SPSS, SAS, and R with the survey package.

With the fall in cost of analysis relative to data collection and integration we should rationally expect more analyses to be done on each data set, and this is indeed the case. Secondary analysis of microdata, especially the publicly-available microdata from US health and social surveys, has exploded. For example, a Google Scholar search with the keywords "*NHANES survival*" lists 20,800 results, and 24,300 results for "*"current population survey" regression*". In fact, the analysis performed on microdata by external researchers may be more varied and complex than that performed in-house by official statistics services.

That is, some of the "High-Quality Statistics Production" addressed by this special issue is, and will continue to be, outside the direct control of the official statistics system, although it relies on the official statistics system for its data and may be paid for by the

[1] Department of Statistics, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand. Email: t.lumley@auckland.ac.nz

same taxpayers. National statistics institutes need to consider cost-effectiveness from the point of view of their mandates, their budgets, and their costs, but governments should look for broader circumstances where it makes sense to fund the national statistics institutes to simplify and encourage external data analysis and reuse.

2. Architecture at National Statistics Institutes

The articles by authors from Statistics New Zealand (SNZ) and Statistics Netherlands (SN) in this special issue emphasize a similar list of needs: information should be collected as few times as possible, administrative data should be integrated with survey data, the data outputs should be flexible, and all this should be done with less money and higher quality. To some extent this is achievable as the payoff from inexpensive computing and accumulated statistical research: cheaper computing allows for higher-level, less efficient specifications of tasks, and improved statistical understanding makes previously disparate tasks look the same. These are the same payoffs that made the R “survey” package feasible (Lumley 2010): modern computing permits the use of an inefficient programming language, and a view of inference based on estimating functions makes programming simpler and more modular (Binder 1983; Estevao et al. 1995).

The SN article emphasizes modularity in the design of the Dutch statistical system. The real issue with modularity is not the concept, but the size of the modules, which typically come in two sizes: too small, and too big. When the modules are too big, the resulting system is inflexible and new modules are always needed; when they are too small, the system is flexible but maintainability and consistency suffer. As the authors say, “*complications become apparent when considering the degree of parameterisation, performance requirements and integration problems between modules.*”

My personal experience with large analytical and data integration systems has mostly been in genomic epidemiology, working with high-throughput genetic data. Flexibility and correctness are overwhelmingly important, while consistency and long-term maintainability are very much secondary. Methodology evolves very rapidly, and it is important to use the best techniques available, but it is relatively less important to maintain analyses from five years earlier. In this context, the analytic system must be small-grained but allow easy assembly of complex tools from simple parts. Nonrelational data stores, together with programming in Java or R, are preferred.

In the official statistics context, consistency and maintainability are primary, and while modern systems are designed to be much more flexible than earlier systems, analysis needs and methodology actually do not change all that fast. This leads to a preference for larger modules, and ones that are assembled in something like a linear fashion, rather than the deeper nesting of a programming language. The SNZ article describes a similar design in their attack on the “Design and Build” bottleneck, where a broad array of modules is linked serially into a “configuration”.

If data are to be reused rather than recollected, a fully linked microdata system is necessary, and a single denormalized data store is attractive. The SNZ article describes such a store, which follows the trend towards increasingly unstructured storage for Big Data. The main risk in such complete data integration is security: as the number of possible data queries increases, it becomes increasingly hard to do any useful content-based screening. Within the national statistics institutes this problem need not be serious, since these organizations already rely on trusted individuals and non-computational data security procedures. Microdata access for outsiders, however, does become more difficult.

When data came as separate surveys it was possible to decide that a specific subset of the variables for a specific survey could be used by a particular class of outsiders under particular conditions. Only a small number of decisions needed to be made, and these could be made in advance. When all analyses rely on data from a comprehensive central store, any combination of variables might be requested and it is much harder to set down rules.

3. Results or Data?

Traditionally, official statistics organizations produced tables for public consumption, and sometimes released subsets of microdata for external analysis. An online analytical server (OAS), as discussed by Krenzke et al. in this special issue, is an intermediate made possible by inexpensive computing and networks. In principle, it could combine the security and much of the simplicity of pre-processed tables with some of the flexibility of microdata analysis.

The difficulty is that decisions about the risk of releasing a particular statistic must be automated. As the authors emphasize, this is not only difficult, but is specific to the particular data set. In contrast to supervised analyses of restricted microdata carried out at internal data centers, it is entirely possible that multiple users could be collaborating to attack the OAS, so that restricting queries based on information available to the same user is not sufficient. Security of an NHIS online analytic server is further complicated by the availability of public-use microdata that could be combined with results from the server.

One of the approaches they evaluate and recommend is to restrict the number of variables available. This obviously reduces the utility of the system, but it is important not just because of the known attacks that they demonstrate, but because of potential unknown attacks. For example, they discuss a situation involving two-way tables of a variable S_0 with each of k variables S_1 to S_k, and where small cells in some of these tables, combined with variable weights, allow an individual to be identified. It is obvious that *this* attack fails if all the cell counts in all the tables are large.

It may seem obvious that this set of tables could not leak *any* identifiable information if all the cell counts were large and the attacker had no detailed information about any of the joint distributions of S_1 to S_k, especially if, say, S_0 is only binary and the other variables are only trinomial. The failure of this conclusion caused widespread consternation in genetic epidemiology in 2008, when it was discovered that knowing S_1 to S_k for an individual in the sample allowed S_0 to be guessed with very high accuracy if k is as large as the typical cell counts (Homer et al. 2008; Church et al. 2009; Im et al. 2012). In survey statistics the number of variables is not high enough for this particular attack to be of concern, but the fact that it came as such a surprise illustrates the problem with high-dimensional data, that our intuitions about sparseness and identifiability become very unreliable.

It may be valuable to explore the possibilities for wider screened-but-unsupervised use of microdata. There would be legal implications, which I cannot comment on, but from the point of view of confidentiality the risk may be quite low. As a comparison, medical research data often contains variables whose disclosure would be genuinely damaging.

When this data is provided under a suitable informed-consent agreement it is not unusual for data access to be provided to other researchers, based on an agreement about how it is to be used, stored, and destroyed. In principle, this is risky. In practice, essentially all disclosures of confidential medical information occur at medical practices or hospitals. Staff who are supposed to disclose this information in response to the right requests are tricked, bribed, or otherwise suborned into giving out the information for the wrong requests (Anderson 2001, p. 166–172). Since epidemiologists have no valid reasons to be giving out personal information, it is harder to convince them to go to substantial lengths to do it for invalid reasons.

4. Economies of Scale and Open Source

The SN architecture explicitly invokes economies of scale, and these are implicit in the NZ design. In these and some of the other articles there is an automatic conclusion that a single system should be large enough for all of the data integration, processing, and analysis for one statistics organization. Some reasons are given for not wanting a system of smaller scope, but there is no discussion about systems of larger scope. If it always makes sense to use the same modular data processing or analysis components within SNZ or within the SN, why does it never make sense to use systems developed externally or cooperatively? In fact, according to the list maintained by the Section on Survey Research Methods, Statistics Netherlands, like Statistics Canada, used to sell and support survey software. Such an approach would also reduce the risk of having no-one available who can maintain or update the system, and would reduce the cost of the necessary testing and audit procedures needed for high-reliability software.

Open source approaches could be one way to gain further economies of scale. In low-level areas such as operating systems and databases there have long been open-source tools of first-rank quality and reliability. Statistical analysis has also benefited from an open-source approach, as exemplified by R (R Core Team 2012), which is a completely open system, and Stata (StataCorp 2012), which is a proprietary system designed to encourage transparent user-written extensions. In many of these systems the basic architecture and implementation are the work of a small, coordinated team, with community development occurring later and providing additional robustness and further features.

If a single national agency, or a small group of them, developed an information management and analysis system (which they have to do anyway) and then made the source code and design available to others, further developments could be made without starting from scratch each time. This "develop then release" approach reduces the burden of coordination at the initial design time and ensures delivery of the initial system (to the extent that any approach ensures this), with the benefits of wider development appearing later.

5. References

Anderson, R.A. (2001). Security Engineering. Hoboken, NJ: John Wiley & Sons.

Binder, D.A. (1983). On the Variances of Asymptotically Normal Estimators from Complex Surveys. International Statistical Review, 51, 279–292.

Church, G., Heeney, C., Hawkins, N., De Vries, J., Boddington, P., Kaye, J., Bobrow, M., and Weir, B. (2009). Public Access to Genome-Wide Data: Five Views on Balancing Research with Privacy and Protection. PLoS Genetics, 5(10), p. e1000665.

Estevao, V., Hidiroglou, M.A., and Särndal, C.-E. (1995). Methodological Principles for a Generalized Estimation System at Statistics Canada. Journal of Official Statistics, 11, 181–204.

Homer, N., Szelinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J., Pearson, J.V., Stephan, D.A., Nelson, S.F., and Craig, D.W. (2008). Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays. PLoS Genetics, 4(8).

Im, H.K., Gamazon, E.R., Nicolae, D.L., and Cox, N.J. (2012). On Sharing Quantitative Trait GWAS Results in an Era of Multiple-Omics Data and the Limits of Genomic Privacy. American Journal of Human Genetics, 90, 591–598.

Lumley, T. (2010). Complex Surveys: A Guide to Analysis Using R. Hoboken, NJ: John Wiley & Sons.

R Core Team (2012). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available at: http://www.R-project.org/

StataCorp (2011). Stata Statistical Software: Release 12. College Station, TX: StataCorp LP.

Journal of Official Statistics, Vol. 29, No. 1, 2013, pp. 177–185

Discussion

Svein Nordbotten[1]

1. Introduction

I would like to thank the Editors for inviting me to comment on articles contributed to this special issue of JOS, Systems and Architecture for High-Quality Statistics Production, and congratulate the authors for their interesting and stimulating articles.

I appreciate that the invitation to the discussants provided the opportunity for "*big picture*" comments on the theme, and shall contribute with comments on system framework, data sources and collection, data organization, storage and computation, dissemination and confidentiality with some references to the submitted articles discussing these subjects. A couple of historic references are provided to illustrate previous work on related topics (Nordbotten 1954, 1975).

2. Framework for Statistical Production Systems

The editor group of John L. Eltinge, Paul P. Biemer and Anders Holmberg has prepared an article in this issue in which a framework is presented to help discussants and readers see how the different topics are interrelated. This framework (denoted FW in the following) illustrates the complexity of modern statistical production systems in which a data set collected for one purpose can be reused for other statistical purposes, how a technological tool can be shared by several parallel processes, how a statistical method or processing algorithm developed for one particular task can be standardized and utilized in a number of other projects, and how extensive the decision and planning tasks become in such systems.

The dynamics of a statistical system are discussed in Section 2.2 of Eltinge et al., and the authors state that *". . .some environmental factors, Z, may vary over time, and thus have an effect on quality, cost and stakeholder utility"*. I would like to elaborate on some factors which I consider extremely important in the context of modern statistical production systems.

There are two types of variations symbolized by elements in Z which should be distinguished. I will call the first type *endogenous* variation caused by feedbacks of previous states of the system itself, but predetermined for the decision maker/planner. I shall denote the second type as *exogenous* variation caused by factors outside the control of the system decision maker/planner. The first kind indicates that in deciding on actions

[1] Professor Emeritus at University of Bergen, Rieber-Mohns veg 1, N-5231 Paradis, Bergen, Norway. Email: svein@nordbotten.com

today, these can have effects on future states of the system at the same time as the effects of previous decisions are inherited by the present planner, while the second kind of variation is caused by factors outside the system and calls for information from outside the system.

Take $U = g_U$ (*Q, C, R, X, Z*), function (1.2.4) of the FW, as a starting point for the design of a separate survey. The function expresses that stakeholder utility *U* of a survey as determined by the vectors *X* and *Z* (*Q*, *C* and *R* are also assumed to be functions of *X* and *Z*), where the elements of *X* denote factors determined by the survey designer and elements of *Z* represent predetermined factors affecting the survey production.

Consider a subvector partition X_W of *X* with elements symbolizing data to be collected and saved. Let the partition Z_W of *Z* denote all historic vectors of this partition X_W, that is be a matrix in which each row is an X_W vector for a previous period

$$Z_W = X_W^{-1}, X_W^{-2}, X_W^{-3} \ldots$$

where the superscripts refer to previous periods.

Obviously, the FW can be used to explicitly take into consideration not only specifications for a survey being considered, but also specifications used for previous surveys. According to $Q = g_Q$ (X, Z, β_Q), functions (1.2.1) in FW, these historic specifications can affect the quality *Q* because historic data increase the efficiency of editing and create a base for time series. They can also affect the cost *C* by symbolizing advantages of lessons learned from previous similar surveys.

Finally, the factors symbolized by Z_W can affect the stakeholder utility. For example, statistics from a considered survey will tend to give more interesting information and a higher utility if statistics from comparable surveys are available. In general, the more statistics are available for a statistical group, the higher the stakeholder utility for this group is expected to be.

In Section 2.2, other dynamic aspects are also discussed as variations in *Z* vector elements due to exogenous factors outside the control of the survey designer. Let the subvector Z_E represent these elements. A typical example of such an element of the Z_E vector is the total survey budget leaving the allocation within the survey to the designer. Other factors represented in the Z_E vector can be methodological and technical innovations, and the expected demand profile for the survey results.

So far the FW has been considered in the context of a single survey. In a modern statistical production system, individual surveys are interrelated and decisions taken for one survey can have an impact on another. Decisions made in other surveys can of course be considered exogenous and represented by elements of the respective *Z* vector by each survey organization, but the interaction will then most likely be lost and the overall design will end up as suboptimal.

As pointed out by Bo Sundgren, the traditional "stove-pipe" organized statistical systems are being replaced in many National Statistical Institutes (NSIs) by integrated statistical systems in which the work on different statistical surveys is closely integrated and interrelations represented in the data collected are taken care of (Sundgren 2010). In such systems, the design decisions for a single survey must always be made with the possible effects on other parts of the total system in mind. A similar point is made by the authors from Statistics Netherlands in their article, in which they write "*In such a situation, one should ideally aim at total network design rather than total survey design. The network is optimised for the same dimensions as is the case when applying total survey design, but the optimisation is global rather than local*".

Let us consider a FW for the whole statistical production system of a NSI consisting of a number of more or less simultaneous activities in different surveys. Assume that the FW reflects such a statistical production system restricted by an available total budget B and a set of general constraints such as limited, shared processing capacity and data archives symbolized in Z. The permissible set of solutions concerning all surveys under consideration will be represented by X vectors satisfying all methodological, system and administrative constraints as well as the cost function $C = g_C\,(X, Z)$. Let budget constraint be symbolized by

$$\sum C_i <= B, \text{ and all } C_i >= 0$$

where $\sum C_i$ is the sum over all allocations to surveys (and other activities) represented by elements of C, associated with respective elements of X. Similar constraints will exist for all shared and limited resources.

The overall task for the NSI managers will be to identify among the permissible vectors the vector X^O which maximizes some function of elements of vector U. The form of this function will express the significance assigned to the respective user groups served by the NSI.

Just as previous decisions will affect the present situation, present decisions will influence future situations. If stakeholder utilities of *future* periods from present design decisions are to be taken into account, future stakeholder utilities must also be evaluated and discounted. Moreover, these kinds of relations can be interpreted and discussed within frameworks similar to that proposed.

Above we changed the view from a framework for a particular survey to a framework for a total NSI production system. Statistical information demands are becoming increasingly global, regional and global cooperation agreements have been made, international statistical organizations have emerged, and the attention may soon be focused on integration of official statistical activities into a *world-wide statistical production system* and the management, trimming and optimizing of such a system. An expanded FW will be a no less important subject for future discussions.

3. Data Sources and Collection

The current outlook for NSIs is characterized by restricted budgets and a growing trend of unwilling respondents to statistical surveys. On the positive side, demand for statistical information and services is reported to be increasing in new groups of users, and technological development of significance for statistical production systems continues to improve and provide access to new or alternative data sources, making the processes more cost effective.

The authors from Statistics New Zealand report in their article that an objective for Statistics New Zealand's ten-year programme for change is "*Maximizing the use of administrative data to reduce respondent load, minimize costs of data collection, and* *maximize data reuse*" and expect that "*By 2020, the organization's aim is that administrative data will be the primary source of information, supplemented where necessary by direct collection*".

If the data are measured in volume, Statistics New Zealand must at the moment rely heavily on sample surveys and have a great saving potential in data from administrative sources. I would guess that in many NSIs, data collected from administrative sources count at least for twice the volume of data collected solely for statistical use. At the same time, the cost of collecting and processing data from an administrative source is less than half of the cost of collecting and processing data per unit from a statistical survey.

The article by the authors from RTI International describes a developed system NIRVANA. The system seems to be designed on the basis of wide experience in survey work and is focused on integration, standardization and use of paradata for improving the quality of the survey results. Particular efforts were made in the project to make access to survey data and paradata for managing the survey operations. From a wider perspective, the NIRVANA project illustrates the importance of working within an architecture in which any survey effort can benefit from experiences from previous surveys.

In most countries, there are still many administrative data sources that have not been exploited by official statistics. In the last couple of decades, communication technology has created new and potential data sources for statistical systems. Statisticians have for some time been aware of credit card and electronic transactions as a potential source of data for statistics. Electronic recording could be a data source for rapid production of continuous short-time transaction statistics and indicators of current economic development by transaction locations. In some well-developed and organized systems, the transaction records could even be linked to records from other sources characterizing the transaction actors, and permitting new, current statistics of great interest for users in trade and industry as well as for economic policy makers. Since the recording and the transfer is done in real time and in general without "response errors", data could be ready for processing immediately and statistics be available on a continuous basis.

A number of other electronic records are being generated that can become important data sources, such as electronic records generated by electronic toll gates using Radio Frequency Identification (RFID), a common sight along the roads in some countries. Vehicles are electronically recorded, location and time stamped when passing the gates, and identified for toll payment by means of a connected car register. Via a car register, the type of vehicle and its attributes including its owner are extracted. These data can be important additions to the ordinary data in preparing statistics on traffic and commuting patterns.

Some countries have a permanent system of registers for such objects as persons, enterprises, real properties, vehicles, and so on within which interrelations between objects in different registers are also currently maintained (UNECE 2007a). In the car register, links to the car owners are kept up to date, in the population register links to the location of owners' dwellings are recorded, and in the enterprise register links to enterprise locations and links to employers are maintained. In the statistical production system, the interrelation links may be used to expand the description of traffic patterns to include owners' socioeconomic or industry groups.

A third fast-growing future source of administrative records, is the electronic tracking already widely used at sea. Most of the shipping fleet today is equipped with Global Positioning Systems (GPS) equipment permitting ships to determine their exact positions at any time. There are systems in operation combining the GPS with electronic position reporting via satellites, radio, mobile phone networks and so forth, permitting the transport patterns to be updated in real time. It is usual to supplement these movement records with size and type of each ship as well as its cargo, and so on. Similar systems have been in place for commercial air traffic for many years. They are now also being installed in cars as a theft precaution. In a few years, all kinds of vehicles will probably be tracked more or less continuously, providing a source for national and international transport statistics.

What kind of methodological challenges do such new sources imply? With a nearly continuous recording of events, a large mass of data will obviously be generated and a scheme for time sampling would probably be required. Since we most certainly will be interested in data permitting analysis of changes, should some kind of panel of objects be designed and observed over time? If so, how frequently should the panel be updated to take into account new objects as well as "deceased" objects (people, cars, ships)? In systems, such as those based on permanent registers, permitting the sampled electronic records to be linked to stored microdata for the objects, challenging methodological questions about the kind of estimation scheme to be applied to obtain the best results will certainly arise.

A second potential source of data is the new social media such as open blogs, Facebook, Twitter, and so on. According to Wikipedia, there are now about 160 million public blogs in existence, of which a large proportion is maintained daily, and every minute of the day 100,000 tweets are sent. Many of these blogs and messages express facts and opinions about public affairs, and they are all open and free to use!

Can important social and economic data be systematically extracted from this stream of blogs and messages and used for the production of useful official statistics? There are blogs and tweets for most kinds of questions debated by politicians and the public in general. In the field of text analysis, a repertoire of methods and computer programs for extracting, interpreting and quantifying text content exists, and can be applied for transforming the contents of textual blogs, tweets, and so on to numerical data for statistical use.

Use of this data source could provide statisticians with data for many of the questions which we currently try to answer by special surveys. We are again challenged by the problem of selecting useful sources from a mass of sources of which a majority probably is of no interest. The solution can be to search for panels of sources proven to keep a given standard of reliability, but new methodological research and development is certainly required.

An interesting consequence of collecting data from administrative sources and continuous streams of data is that the traditional strong relationship in time between survey collection and processing of data becomes less important. Some of the peak workloads associated with censuses have already disappeared in countries producing census statistics from administrative data. Data will be collected and saved to serve future needs, and processing may be carried out when needed, relying on already collected and saved data.

4. Data Organization, Storage, and Computation

In many statistical organizations, after being processed and used for primary publication microdata are saved in some kind of statistical archive (warehouse, database) for reuse in some future connection which can be specified by a sub-vector Z_W of the Z in FW. By saving microdata for future use, we can obtain a higher future performance because more microdata are available for each object of the population, permitting a wider set of tabulations and the creation of more realistic analytical models.

It will be useful to distinguish between microdata and macrodata stored in the statistical archive. A cube or OLAP-approach, as adopted by Statistics NZ and many other NSIs, has proved to be a good strategy for storing macrodata and serving requests for special statistics. Systematic storage for microdata is, however, needed for responding to requests which cannot be served by the cube storage or for serving researchers' requests for access to microdata. Such user needs can be served by a cumulating system of cleaned microdata files on which it is possible to remove unwanted records, link object records from different files, directly and by reference, and to store the results as a new generated microdata file ready to serve future requests. The cost of such system readiness is data redundancy, since the same data will appear in more than one file. An alternative approach, considered and discussed for decades, is to record each atomic data element and retrieve the data for each request (Sundgren 1973); the "single-cell" approach of Statistics NZ seems to be a partial realization of Sundgren's model.

For effective macrodata and microdata storage and retrieval, a well-developed metadata system is crucial. It must be automatically maintained and updated for each transaction/operation on data in storage. Efficient reuse of data is completely dependent on a well-structured metadata system from which users can obtain information to form their requests. We can think about representation of metadata system effectiveness in the FW as symbolized by an element of the Z_W vector indicating the coverage of the metadata system in the past, while a corresponding element of the X_W subvector expresses a planned improvement of the metadata system.

There is another important issue linked to integrating microdata. The more data you have collected about an object, the easier it will in general be to identify errors in the object data. This creates a dilemma: What to do when editing of newly collected data for an object reveals errors in old data included in the archive and already used in several productions? If the NSI policy is not to correct microdata in the archive because of desired consistency between earlier published statistics and the archived microdata, the producer will have to ignore errors in previously released statistics even though new statistics then may not be consistent with statistics previously published. On the other hand, if the NSI policy is to improve the quality of all statistics released, the producer must correct data in the archive, rerun all productions based on the corrected data set, publish correction messages and notify all users who have received statistical services based on the erroneous data. My guess is that the latter policy in the long run is unrealistic, and that NSIs must admit that errors always can appear in statistics. A compromise will be to make corrections in data less than N years old, and keep the newer data files in a preliminary storage before transfer to the archive.

In the Nordic and some other countries, the introduction and maintenance of official central registers for population, enterprises, dwellings/properties, and so on including unique and permanent object identifiers, have been a political goal in order to reduce inconvenience to the public and government registration efforts, and save expenses. The official registers are also in some cases available for private use, for example by banks and insurance companies. This fact makes data from administrative sources particularly useful for the NSIs, since the data records received from administrative organizations are pre-identified and data from different sources can be linked as if they originated from a huge single source. Because of the permanence of the identifiers, records from different years or periods can be linked and microdata time series created from all contributing sources (Nordbotten 2010).

Except for estimation methods used in the context of sample surveys, the computation of statistics has been dominated by simple aggregation and has remained unchanged for centuries. An important discussion is going on among survey statisticians about design-based versus model-based statistics. Is the time due for a general paradigm shift? In statistical processes like data collection, editing and imputation, attribute values are frequently considered realizations of stochastic variables. Why not consider any collected data set, for example a population census, as a random sample of a superpopulation and use the huge amount of existing paradata to develop quality measures for the estimates published? In such a scenario, a cell-based storage structure as outlined by Statistics NZ can be interesting component.

5. Dissemination and Confidentiality

The objective of official statistical production systems is to provide, with their allocated resources, statistical information requested by a wide set of user groups. Traditionally, this objective was fulfilled by publishing statistical tables, frequently designed before data were collected and based on the needs expressed by user groups dominated by state authorities. In more recent times, statistical information systems have been regarded a common service to users from all organizations, industries, education, research, media and the general public by responding to special requests.

The dissemination service can be provided in different ways, as discussed by Statistics NZ. There are, however, two aspects, *instant online service* and *access to microdata*, which are of particular interest in the context of modern statistical production system architecture (UNECE 2007b). They have in common that both create disclosure concerns.

In their article, the authors from Westat/NCHS discuss challenges in planning and preparations for a web-based Online Analytical System (OAS). The objective of the OAS is to serve researchers and other users with special needs with data from the National Center for Health Statistics (NCHS). The authors are well aware of the problem of balancing user service and disclosure risk. To provide the data requested, the data have to pass the Statistical Disclosure Control (SDC) treatments. How to control the risk for disclosure subject to the anticipated attack strategies of an intruder are discussed and solutions outlined. Even though the discussion is limited to the survey environment of the NCHS, the discussion also can be instructive for statisticians working under more general conditions.

It can be interesting to consider disclosure in the context of the FW. Let disclosure risk for a specified type of microdata be defined as

$$C_d = p_d \times D_d$$

where risk C_d can be regarded as a cost element in vector C of the system, p_d is the probability for disclosure of the considered microdata and D_d is the damage suffered by the concerned physical person(s) or owner(s) of a harmed object due to this disclosure. Reducing C_d means usually restricting the available statistical information (reducing U, P, Q) and increasing the treatment cost, another element of C. A discussion of the disclosure risk in terms of objective estimates for the probability of disclosure and the size of the damage if the data are disclosed is rarely seen and would be useful.

How to provide access for researchers (and why not also other users?) to microdata files without disclosing confidential data? This has become an extremely important question when real-time online service systems are being considered. Several approaches, including removal of all identifying elements, distorting data by random errors, authorization arrangements, and so on, have been proposed and tried. They have all their pros and cons.

One strategy can be that the online researchers (users) are given access to "invisible" microdata files, that is, are restricted to seeing shadow views of the relevant microdata files. In shadow views, the viewer cannot see any attribute data, but whether attribute values are present or not, whether a value was observed or imputed, and so forth. The users can be permitted to create new microdata files by operating on existing files, form new files by integrating existing files, study their shadows, and carry out analytic computations on data based on a repertoire of approved analytical programs. The disclosure problems of researchers' online reuse of microdata would then be limited to the screening of analytical results and be similar to those discussed in the article by Westat/NCHS, and the same SDC treatment could be relevant.

6. A Final Comment

Any architecture for statistical production systems has to be designed in accordance with statistical (and other relevant) legislation. Several ideas discussed above would in some countries require modifications of existing statistical laws. Some NSIs have the advantage of one or more official identification systems. Do they have the required basis for legal access to administrative data from all organizations using the systems? Are the legal basis and the conditions for an NSI to store and integrate microdata for an indefinite time period satisfactory? Is the legislation safeguarding a confidential transfer and treatment of microdata up to date? Does the legislation permitting access to microdata reflect a fair balance between the public needs for information and required confidentiality? Statistical laws need periodic modifications to satisfy social and economic development and represent a significant premise for the statistical system architecture. Statistical legislation should therefore be a part of future discussions of statistical system architecture frameworks.

7. References

Nordbotten, S. (1954). Theory of Production and Production of Statistics, Translation of a paper Produksjonsteori og statistikkproduksjon (in Norwegian) published in Stimulator, 7, 28–31. Available at: http://nordbotten.com/articles (accessed March 2013).

Nordbotten, S. (1975). On the Design of Statistical Information Systems. In Systemering 75, Festskrift till Børje Langefors, J. Bubenko and M. Lundeberg (Eds), Lund: Studentlitteratur. Available at: http://nordbotten.com/articles (accessed March 2013).

Nordbotten, S. (2010). The Statistical Archive System 1960–2010: A Summary. Nordisk Statistikermøde i København 11.–14. august 2010. København. Available at: http://nordbotten.com/articles/ (accessed March 2013).

Sundgren, B. (1973). An Infological Approach to Data Bases. Stockholm University and Statistics Sweden, Stockholm. Available at: http://sites.google.com/site/bosundgren/my-life/AnInfologicalApproachtoDataBases.pdf (accessed March 2013).

Sundgren, B. (2010). The Systems Approach to Official Statistics, Chapter 18 in Official Statistics in Honour of Daniel Thorburn, 225–260. Available at: http://officialstatistics.wordpress.com (accessed March 2013).

UNECE (2007a). Register-Based Statistics in the Nordic Countries. Review of Best Practices with Focus on Population and Social Statistics. United Nations Economic Commission for Europe. Available at: http://www.unece.org/stats/publications/Register_based_statistics_in_Nordic_countries.pdf (accessed March 2013).

UNECE (2007b). Managing Statistical Confidentiality & Microdata Access. Principles and Guidelines for Good Practice. New York and Geneva: United Nations.

Journal of Official Statistics, Vol. 29, No. 1, 2013, pp. 187–192

Discussion

Stephen Penneck[1]

1. Introduction

The redesign of statistical systems to reflect an agency-wide strategic architecture is one of the most fundamental challenges a statistics agency faces. It is both necessary and difficult.

It is necessary because it will yield the cost savings and quality improvements that funders and users demand. The development of methods, systems and processes to common standards under a unified architecture yields savings in development and maintenance costs, which are timely given ever tightening budget constraints. At the same time common metadata standards, methods and approaches guarantee quality standards at a time when users become more demanding. However, although the agency-wide benefits of this approach are self evident, statistical producers have found this approach hard to deliver.

The strategic redesign of statistical systems poses a number of difficulties and the articles in this issue give some good examples of these and how statistical producers have attempted to overcome them. The articles also underline the importance of statistical leadership, and my discussant note focuses on this aspect.

2. The Strategic Challenge

The extent of the challenge of redesigning systems will vary for each statistical producer, depending on its history of statistical development and the extent to which past development has been at a strategic level. The challenge is particularly daunting for producers that have a diverse infrastructure. Diversification may have arisen for a variety of reasons. One cause, as in the Office for National Statistics (ONS) in the UK, is the move from a largely decentralised statistical system to greater centralisation. The component agencies in decentralised systems tend to follow separate and diverse IT and methodological approaches. As each production area merges into the centre, it takes with it its own particular approaches. Or it may result from the agency maintaining diverse systems in different geographical areas. This is the background set out in the Nealon and Gleaton article, which describes the National Agricultural Statistics Service (NASS) experience (and I shall refer to as the NASS article). Or, as in the article by Seyb, McKenzie and Skerrett from Statistics New Zealand (the SNZ article), diversification may arise from a process culture (also a factor at ONS in the UK, I think), which together with a hierarchical organisation structure led to investment decisions being driven by the need to replace ageing production systems in production silos. At RTI International (the article by

[1] President, International Association for Official Statistics, Former Director General, Office for National Statistics (ONS), UK. Email: stephen.penneck@gmail.com

Thalji et al.) it appears to have arisen from a project approach, designed to meet specific requests from users. In addition, diversification may arise on the technology front as a result of the preference in some agencies two or three decades ago for IT services to be bedded out to specific statistical areas, as described in the article by Struijs et al. (the CBS Netherlands article). This local ownership of IT weakens the strategic drive and often leads to a multiplicity of different solutions and software licencing arrangements.

Whatever the reason for the problem, I think we would all agree that the costs of maintaining such diverse systems escalate. It is useful to see this documented in the SNZ article, and there are also good examples in the NASS article. The CBS article sets out their Enterprise Architecture approach, being the framework needed to successively redesign the statistical processes of an organisation. But bringing systems and methods under a common architecture can be very difficult where the cultural tradition favours diversity. Local statistical solutions for individual business areas, be they national accounts, prices or population statistics, will usually produce optimal solutions for those areas. For example, software that supports a particular editing approach which suits price data may differ from that which suits social surveys. Users, to the extent that they have a view on statistical methods, will support local solutions, as will the statisticians that manage those systems, who may have had a key role in their development. The benefits of locally focused development will be very evident at the local level. Strategic solutions are often suboptimal locally, and their benefits are apparent only for the agency as a whole. In the end, the issue may come to a head if there are simply not the resources to continue with locally led development. However, managers of local statistical systems will have invested much of their time and energy in developing and maintaining those systems, and will have considerable expertise in them, and not necessarily be expert in the strategic solutions being proposed.

3. The Role of Statistical Leadership

In these circumstances, the need for effective top level statistical leadership is clear. By statistical leadership I mean the ability of the senior professional statisticians in the organisation, working as a team, to meet the strategic challenge, set the direction and take their stakeholders with them. Statistical leaders can do this by propounding the benefits of strategic change to all stakeholders, but especially to funders, to users, and to staff.

Strategic development of statistical systems is expensive and cannot always be funded internally. External funders will principally be concerned about value for money: that is, control of costs and delivering financial benefits. Where an organisation is primarily funded through users, for example, by charging for surveys, judgements need to be made about the extent to which overhead charges can pay for statistical developments. Government agencies which are mainly vote funded by government may need to bid for development funds. The strategic nature of statistical system development generally leads to a statistical programme made up of many projects. These will often include technology projects, projects to determine best methods, projects to redesign business processes and projects to deliver production systems. Indeed technology projects are often prerequisites for statistical transformation, which requires common technology platforms and standards. Thus the NASS article describes a project to consolidate network services across their Field Offices; standardise and integrate survey data and metadata; and generalise survey applications and databases. The SNZ article covers a modernised IT system, standardised processes and methodology. The RTI article covers a centralised database, case management system, survey instruments and a quality evaluation system. The CBS article notes that the major financial savings will often not be apparent until the programme is complete, although some articles note the possibility of realising some savings along the way.

Each project is complex and generally there is a high degree of interrelationship between them, making the estimation of the overall costs uncertain. In addition, financial benefits will often not be fully apparent until production systems are delivered, often much later than the early infrastructure development projects. At the beginning of the programme business cases will be produced, highly uncertain given the nature of these developments, but essential to secure the funding needed. Unless funders are made aware of the complexity of the developments planned and the attendant pitfalls, there is a high risk that failure to deliver the benefits in the business case will adversely affect the reputation of the agency. The SNZ article stresses the expectations that public sector agencies have on them to deliver cost-effective statistical systems.

Leaders will also need to ensure that users fully understand the strategic benefits. Locally managed systems will often deliver rapid responses to changing user requirements. Users may be concerned that a strategic solution will be less responsive. They may also be concerned that a significant change of approach will lead to discontinuities. Users will need to be persuaded of the broader benefits of the new approach and that the existing approach may not be sustainable in the long run.

The agency's staff is the key stakeholder group. The new development approach will require a high level of detailed understanding by all staff. It is likely to require a large number of projects, all to operate within a common architectural framework. Project managers and business area managers will need to understand the new approach and be able to see the implications for their work and how the approach can best be used. Staff in the agency will often be alert to the benefits of change. It is interesting that in the NASS article staff were asked where savings in the future might come from, and they identified the more effective use of staff, seen in the wasteful duplication of functions in the agency.

Agencies often use statements of principles to articulate the new approach. The NASS article uses a set of four criteria: centralisation, integration, standardisation and flexibility to provide direction to its transformation. SNZ defined a set of nine "guiding principles". CBS founded their enterprise model on a set of key principles. RTI identified standardisation, integration, and data access. Such statements of principle are powerful ways of communicating the strategic approach to staff. The degree of commonality in these sets of principles in different organisations is striking.

With each of these stakeholder groups – funders, users and staff – it is for the leadership in the agency to articulate the vision, and set out the benefits in ways that can be readily understood. A relentless focus on communications is essential, together with setting up good governance and evaluation approaches. The NASS article pays tribute to the NASS Administrator's support and the role of the senior executive team. There will always be differences of view amongst senior executives, especially where a development programme has led to some re-ordering of priorities. Staff will be looking for the wholehearted commitment of the senior team to the changes that are about to be made, and a divided team will have a considerable impact on how well the programme can be delivered. SNZ recognise the need for a culture and change leadership programme to develop strong leaders and develop a performance culture. The CBS article also notes that insufficient communications in the early stages led to misunderstandings.

4. The Complexity of Statistical Development: The Need for Methodologists and Technologists to Work Together

The need to upgrade and develop technical architecture and the need to improve statistical methods often go hand in hand, and technologists and methodologists need to work closely together to ensure success. Both NASS and SNZ describe a move away from traditional transactions processing, with its emphasis on speed and repeatability of process, towards analytical processing, providing more flexible and powerful analysis. Technology innovations such as thin client architecture, virtualisation of servers and the development of Agile development approaches, as described in the NASS article, lend themselves both to the easier consolidation of statistical systems and the active engagement of statisticians in their development.

Although we have come to view a generalisation of the statistical process as set out in UNECE (2009) as the norm, that high level view covers a degree of complexity, leading to what SNZ describe as the "design and build" bottleneck. At each stage of the model there are options and variations, often tied up with particular software choices. In addition, the detailed coding required in software solutions, even for proprietary software, can be complex and require a high degree of statistical understanding. Close collaboration is required between the designers (traditionally the statistical methodologists) and the system builders (traditionally information technologists). A third player in the development triangle is the business owner of the system being developed, often a statistician, but possibly from a related analytical discipline, such as an economist or social researcher, or maybe an expert process manager. The SNZ article describes this group as "subject matter specialists". As the NASS article makes clear, one of the most frequent causes of failure is a failure to meet business requirements, caused by a lack of understanding between the three parties concerned. Both NASS and SNZ have employed the agile approach to help build collaboration by requiring constant communication between the parties. In fact SNZ have gone further. By developing a highly flexible statistical infrastructure and a generic set of tools, they have empowered their subject matter specialists enabling them to develop systems themselves, and providing a quick and low-cost development approach.

Collaboration between professionals builds over time and as the transformation progresses through a number of projects there is a degree of organisational and individual learning. Using external IT providers or outsourced staff can be a barrier to this, although it may be necessary if sufficient skills – either specific skills or the quantity of skills – are not available in-house. In most countries, contractors will rarely have experience of developing statistical systems, as such systems are not widely used outside statistical agencies. Consultants tend to underestimate the complexity of these systems, and if they are only contracted for a single project, do not have the opportunity to learn. A longer term strategic relationship is needed, but can be difficult to sustain, given the agencies' need for contestability and the contractors' uncertainty about the value of long-term skills investment. The employment of contractors for statistical development is difficult to envisage in the "analyst builds" approach of SNZ. RTI looked at a variety of commercial off-the-shelf packages for its data collection project, but decided to develop in-house to maximise its design control.

How can we manage such complexity? Good project management techniques are essential, together with programme management to ensure interdependencies between projects are understood and managed. SNZ has developed an approach to complexity by breaking down the transformation into a number of phases, and setting targets for each phase. Furthermore, they have used a prototyping approach with their scenario store: this essentially mirrors the live production environment. Statistical leaders also need to think about the organisational implications of enterprise-wide strategic developments. The CBS article describes the need to set up new units to undertake new corporate functions (e.g., their Data Service Centre), while in the UK, the ONS has brought together methodologists and IT experts into a Strategy and Standards Directorate, providing an innovative focus for the Office, and ensuring that their developments are aligned with the corporate strategy and standards.

5. The Need for Strong Governance

The NASS article is right to point to the need for strong governance as part of the answer to complexity. It can help to ensure that the strategic benefits are realised, and that business requirements are met. As noted earlier, all articles have a set of guiding principles which describes their visions.

The NASS article identifies the role of a Senior Project Manager, which in the ONS we called an Executive Sponsor, or Senior Responsible Owner: a board level executive who takes on the role of ensuring strategic alignment and that the agency level business benefits are delivered. SNZ have set up Portfolio Committees. RTI had a team of senior executives as a steering committee.

The NASS article also describes an Enterprise Architecture Council (in ONS known as the Design Authority) that has the role of approving technical standards and of ensuring that projects meet them. SNZ have a similar Standards Review Board. Without some such Authority it is very easy for the results of such developments to be an updated technology with little change to business process or methods, so that the overall business benefits are not delivered. The CBS have created an Enterprise Architecture Board to interpret their strategic goals and make decisions on how to operationalise them.

SNZ recognise that their development model, using analysts as builders, requires a high degree of control to prevent the problems of the past resurfacing. Hence they build limitations and controls into what the analysts can do.

6. The Need for Organisational Learning

A number of agencies have now been working on strategic redesign for some years. The CBS has been following its current enterprise architecture approach for about seven years. SNZ has had more than one attempt at standardising its statistical production. The ONS has also been on this path for a while now (see Penneck 2009), using different approaches. There are some similarities here, with the agencies moving from a large-scale infrastructure development approach to smaller-scale rapid developments.

The RTI has a well-developed approach to evaluating the success of its developments, which relates performance to success criteria based on its original principles. RTI also plans to monitor development costs in the new strategic environment compared with ad hoc developments. The other articles say less on how they will evaluate their programmes, although they are clear about the need to measure benefits.

Most of the articles set out the points they have learned along the way, and bringing them together will be a valuable aid for all agencies engaged on this work. The work described in these articles is necessarily long term, and the membership of project teams and the holders of key posts will change over time. The leaders of an organisation will want to ensure that the lessons derived – the positive as well as the negative – are learned by the whole of the organisation and not only by a few individuals who may move on (or worse, only by external consultants). Organisational learning is the key to organisational improvement, and can only be successful if the lessons are widely disseminated within that organisation.

Is there scope for greater international collaboration? The answer must be "Yes". Reading RTI's development of a case management system reminded me of ONS' similar attempt some years ago. Certainly statistical producers should be encouraged to write up and publish their experiences in this area: there is a great deal of common ground and potential for international learning.

7. Conclusion

In this short note I have attempted to draw on some of the experiences set out in some of the articles in this issue, together with my own experience in ONS, to underline the importance of statistical leadership for an organisation which is embarking on strategic redesign of its statistical systems. By exercising statistical leadership, the senior statisticians in the organisation can respond to the strategic challenges of a change typified by:

- a legacy of diverse systems, methods and processes.
- benefits most evident at a corporate level.
- a high degree of complexity.
- a need for good governance.
- a need for organisational learning.

8. References

Penneck, S.J. (2009). The ONS Statistical Modernisation Programme What went right? What did we learn? Paper given at the Conference on Modernisation of Statistics Production. Available at: http://www.scb.se/Grupp/Produkter_Tjanster/Kurser/ModernisationWorkshop/final_papers/D_1_management_Penneck_final.pdf (accessed March, 2013)

United Nations Economic Commission for Europe (UNECE) (2009). The Generic Statistical Business Process Model. Available at: http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model (accessed March, 2013).

Journal of Official Statistics, Vol. 29, No. 1, 2013, pp. 193–201

Discussion

Claude Poirier[1], Claude Julien[1], and Robert McLellan[1]

1. Introduction

National Statistical Offices (NSOs) must continuously evolve to produce the broad range of information required by law, regulations and numerous users of official statistics. Typically, NSOs have redesigned, enhanced, added or dropped production lines and survey processes that are executed fairly independently from one another. The three articles presented by NSOs in this special issue clearly show that they are facing similar budget constraints, which translate into a set of common drivers for the transformation of their statistical programs and organization. Amongst them, we note the desire to ensure cost efficiency in the processes; to reduce response burden, especially for businesses; to improve coherence; to maintain high quality standards; and to improve the responsiveness of statistical programs.

This discussion reviews the submissions from three NSOs who have moved forward on their path of modernization. Statistics Netherlands writes about its adoption of an "enterprise architecture" approach through process standardization and focus on design and data "steady-states". Statistics New Zealand writes about its Statistics 2020 initiative and the key actions they have taken to transform their statistical production systems, with important changes to their corporate design and database design. Finally, the National Agricultural Statistics Service (NASS) in the United States shares its experiences of moving from a highly distributed approach to centralized infrastructure and data management.

It is clear from reading the three articles that, in order to meet these objectives, redesigning and enhancing current independent production lines or processes is not sufficient. NSOs require much larger scale transformation in how they produce and deliver information and how they organize themselves accordingly. This discussion outlines the main solutions proposed by the articles' authors in Section 2 and how they were implemented in their respective organizations in Section 3.

2. Solutions

National statistical organizations are all facing similar issues with respect to efficiency and risk management. At the same time, they have access to technological and methodological advancements. Several good solutions are proposed to react to budget constraints, to

[1] Statistics Canada, 150 Tunney's Pasture Driveway, Ottawa, ON, K1A 0T6, Canada. Emails: Claude.Poirier@statcan.gc.ca, Claude.Julien@statcan.gc.ca, and Robert.McLellan@statcan.gc.ca

improve productivity, to reach unprecedented coherence, to reduce risks of errors, if not simply to adopt common business processes. Most of these solutions address several goals, as described below.

2.1. Centralized Databases

Centralized databases provide a number of benefits to a statistical organization. They allow access to a vast amount of survey and administrative data that are at the core of the cost-efficient transformations taking place in the NSOs. Such sharing of data avoids the collection of similar information, if not simply to serve as proxy values for designing surveys or modeling data.

Centralized databases are the result of good information management as long as they are interpretable through appropriate documentation and detailed metadata. In their articles, the three NSOs mention that the various versions of their data – or audit trails – are stored in their centralized data warehouses, in order to allow rerunning or evaluating any process. Other benefits include a less confusing environment for processing purposes, a more efficient use of all information during the survey cycles, a potential increase of NSOs' responsiveness, and an input to corporate memory.

However, centralizing data requires more complex management of data confidentiality within the organization. Indeed, many survey areas collect information that should not be shared with other areas. Therefore, centralization and the related access control bring a risk of mismanaging the accessibility that may result in undesired disclosure.

Users of administrative data often face issues with their timeliness. Data modeling should be considered as a solution to this challenge. Statistics New Zealand plans on using modeling as an opportunity to modify tax data that do not correspond exactly to the variables of interest. This approach can be expanded to introduce trend adjustments to any administrative data when they do not fit the appropriate reference periods, or to enable the exclusion of small units during the regular data collection and allow for adjustments through models.

2.2. Standardization

Standardization is a key principle for most NSOs facing financial pressures. Statistics New Zealand targets standardized processes, methods, tools and systems. This will help them to reduce IT maintenance costs and risks. NASS addresses standardization through centralization and consolidation. Statistics Netherlands invests in standardization to enhance the flexibility of statistical processes, to increase their uses and to reduce costs.

Standardization means that common approaches are used to perform business functions. Standardization usually requires the consolidation of some systems to maintain a reduced set of functions. In this context, generalization is carried out to make sure that the reduced set is developed with an acceptable level of flexibility that still satisfies a wide range of applications. This approach targets reusable building blocks for which behavior is controlled with parameters. To be usable in multiple instances, large monolithic solutions need to be broken into smaller components. This concept of breaking large solutions into smaller components is central to the 'plug & play' architecture being pursued among NSOs (HLG-BAS 2012).

Statistics Canada is currently modernizing its Corporate Business Architecture (CBA). Key elements of this involve creating standardized collection processes, tools, and data that feed downstream social, economic, and Census processing. Standard processing platforms for economic and social statistics are also being developed. Since processing tools must be able to interact, it is necessary to standardize interfaces and data. For that reason, Statistics Canada has had in place a central metadata management database, which provides a standardized metadata space covering important aspects of statistical production.

Statistics Canada has adopted the Generic Statistical Business Process Model and has contributed to the Generic Statistical Information Model (GSIM). It is our belief that GSIM will provide a basis for our information standardization activities as part of our CBA modernization.

We realize that standardization should not interfere with extensibility. For instance, while classification is part of standardization, we have found that users frequently extend classifications in use for valid business reasons. In our experience, expecting there to be no refinement of standards is likely unworkable. Establishing clear working mechanisms to allow for controlled extensibility with associated mappings is a reasonable compromise to support business innovation.

2.3. Integration

Integration is the connection of various centralized databases, functions and standard tools to produce an end-to-end production line. NASS uses integration as one of four criteria to determine if a transformation met their objectives. Statistics New Zealand combines a set of processes into a configuration that is applied to a set of data. Statistics Netherlands goes further with an integral approach that optimizes a set of production lines that may share the same source of data.

Statistics Canada's Integrated Business Statistics Program (Ravindra 2012) is somewhere between the latter models. When completed, it will be a large-scale platform that will support the production and analysis of more than 100 business surveys and make extensive use of tax data. It will use generalized systems for sampling, edit and imputation, estimation and time series, and other standard systems for collection, analysis and dissemination. The single platform will require much less IT maintenance than would individual production lines. However, considerable IT resources are required to program the numerous data manipulations to integrate the standard tools into the single platform.

2.4. Flexibility

The NSOs demonstrate a common interest for a flexible environment. As mentioned by MIT-ESD (2001): "Flexibility may indicate the ease of 'programming' the system to achieve a variety of functions. It can also indicate the ease of changing the system's requirements with a relatively small increase in complexity and rework." In other words, flexibility introduces the concepts of (a) adaptability and (b) extensibility.

(a) Concerning adaptability, the idea behind most transformation projects is to standardize processes and achieve systems that can be adapted to most applications. It must be noted that such an effort rarely provides the level of flexibility that the initial environment offered, since flexibility competes with standardization. This is the price to pay in the quest for more efficient processes. Senior management should make sure that business cases address this "global optimum vs. local optimum" trade-off early in the transformation project.

(b) With respect to extensibility, it is important to set up processes that can be modified in order to satisfy evolving requirements. This is a prerequisite for a responsive architecture. This aspect of flexibility is addressed only implicitly by the NSOs, but Eltinge, Biemer and Holmberg address it explicitly in their proposed framework, through the dynamic features of quality, cost, risk and relevance. They purposely suggest that performance criteria should go beyond traditional metrics like bias and variance.

2.5. IT Strategies

The approaches outlined by the authors vary in the specificity of the underlying information technology and solutions. Central to the approach taken by NASS is the creation of a centralized infrastructure function in line with modern IT practices. The modernization of networks, centralization of servers, use of both desktop and server virtualization, and deployment of standardized desktops are all part of mainstream IT modernization activities in industry today. Both Statistics New Zealand and NASS report on important changes in their approach to providing databases, with Statistics New Zealand highlighting a shift to Online Analytical Processing (OLAP) approaches away from Online Transactional Processing (OLTP). The adoption of more "storage-hungry" solutions is offset by the use of modern storage technology in flexible storage solutions. Similarly, processing power to support a variety of approaches enables them to create a richer data and metadata environment with powerful flexibility. The importance of a comprehensive enterprise architecture which incorporates the IT aspect reflects the need to bring IT and non-IT communities closer together in a collaborative setting.

3. Good Practices

The transformation of a statistical production system can only be successfully implemented using sound practices. This section outlines good practices that are reused by NSOs. While these objectives address current risks, they may introduce new risks that should not be ignored. Special attention is paid to these new risks below.

3.1. Clear Objectives

Transformation means changing the way we do things with the objective of improving the process and reducing vulnerabilities. Transformation projects always start with a business case to demonstrate benefits of proposed solutions and list clear objectives. On that matter, Statistics New Zealand, Statistics Netherlands and NASS initiated their overall projects with well-defined objectives. In-scope and out-of-scope processes are clearly listed, not only to drive the governance but also to avoid misinterpretation by affected employees.

Such projects generally try to streamline processes within a new architecture where tasks and activities are centralized and standardized. The several processes involved increase the impact that any failures would have on the overall success. This translates into a new risk: the interdependency of transformed processes. We would then recommend keeping transformation tasks as small as possible and spreading these into realistic timelines to avoid overloaded schedules.

The determination of priorities comes after the objectives. It was noted that most NSOs adopted a reactive approach in motivating priorities based on obsolescence of processes and systems. We believe that a more proactive strategy should be considered. A long-term development planning (say over ten years) could be initiated, and then updated on an annual basis. This would ensure a continuous maintenance of the statistical programs. It would also avoid surprises in the budgeting process, and would serve as input for a human resource plan. Statistics New Zealand provides a good example with such a long horizon.

Employee involvement or buy-in is essential to the success of the transformations. The articles offer different approaches, that is, top-down or bottom-up. From our experiences, miscommunication may emerge when employees do not understand the pressure the organization is facing. Reluctance to change may also be observed while ideas may contradict each other. We would then recommend planning for an appropriate combination of top-down and bottom-up approaches. Senior management would describe the indisputable pressures facing the organization, consult end users, identify areas to be improved, set targets and action plans, invest in a communication strategy, and then listen to employees' concerns while looking for innovative ideas on their part on how to implement and optimize a transformation.

3.2. *Governance*

Governance is probably the most important aspect of large projects. Without governance, the project scope, schedule and budget would not be monitored closely enough to ensure the success of the project. Statistics Netherlands describes a well-structured governance model that involves a board of directors, a portfolio board, an architecture board, and steering committees dedicated to each project. Statistics New Zealand implemented a very similar structure with a transformation team, a standards review board, portfolio committees, and steering committees. While NASS also has a similar structure, they go beyond the Statistics Netherlands and Statistics New Zealand model by setting a change control board to govern the transition to the new architecture – an exemplary approach.

Statistics Canada has put in place a similar governance structure to those demonstrated by the NSOs. The governance starts with a corporate management framework, that is, a vision that lays out the mandate and objectives of the organization and the related vulnerabilities. It brings together management principles to enable financial and operational success as well as to monitor the interdependency of projects.

Statistical organizations would benefit from a project management office, on site, to assist senior management and project managers within the governance structure. Such an office would offer advice and tools to guide project managers in their activities and assist them in documenting their projects throughout their phases.

3.3. *Cost-Benefit Analysis*

All articles in this special issue describe cost-benefit or business case analyses; some more explicitly than others. Statistics Netherlands states it very well: "The goal of cost-efficiency is only reached if cost-benefit considerations are correctly taken into account when taking decisions, of course. We look at cost-benefit considerations first for the architecture, then for deciding on standardization and finally for project decisions on individual (re)design projects in the context of portfolio management."

Cost should not only be measured financially but also in terms of workload, expertise required, and capacity to manage large and complex projects. In the articles, NSOs do not formally address transition costs from the former to the new architecture. This requires considerable effort, especially in migrating applications, training staff with new systems, and putting in place support units. We would recommend neither to underestimate transition costs nor ongoing maintenance costs when planning transformations.

The transformation projects described by the NSOs are all driven by financial pressures, but to varying degrees. Pressures to do more with less have been around for many years, but particularly so recently; it seems likely to remain this way in the years to come. However, financial pressures should not be the only driver for transformation. It should be initiated because and when NSOs can, not when NSOs must. Statistics Netherlands' strategic decision to transform processes is a good example of this. That said, a portion of any savings harvested from a transformation should be reinvested in research and development or other transformation initiatives to further improve the efficiency or the quality and quantity of information delivered by the NSOs.

3.4. Statistical Tool Box

The development of a tool box that offers key statistical functions is also a good practice suggested by the NSOs. Whether the tools are standardized, generalized, centralized, consolidated, or common, these tools are to be used by most statistical processes within organizations. A tool box is a way to reduce duplication of functions, to help guarantee their sustainability, to ease the transition of staff across surveys, to reduce implementation errors, to focus research efforts, and to ensure sound methods.

Any tool box must be maintained. The supported functions must stay in line with the survey requirements. Therefore, resources must be set aside to ensure the relevance of tool box components. This means research activities must be funded to keep up with state-of-the-art methods, as alluded to in Section 3.7. Tools must be classified as emerging, recommended, phasing-out or obsolete, and their renewal should be included in the organization's long-term plans. Furthermore, governance must prescribe mandatory tools and have exceptions considered, approved and monitored by corporate review boards.

Given all its positive aspects, a tool box is not free of issues. For the employees, a tool box may be perceived as an obstacle to their creativity. In the envisioned environment, statisticians should dedicate their valuable time to making good and innovative uses of standardized systems and conducting research into improving these systems, rather than redeveloping customized systems that happen to be fairly similar to each other. For other employees, there is a risk of the press-button syndrome where statisticians simply run systems with a set of parameters without proper theoretical knowledge on how the system works and critical look at what the system produces. Such behavior may generate costly errors, which could include revising estimates and could contribute to a reduction of end user trust. This should be avoided by ensuring that staff have the appropriate training and knowledge to run these systems effectively.

3.5. Data and Metadata Management

Data and metadata need to be consistently and carefully managed, and appropriate decisions must be made upon the known quality of a particular dataset. Automation of processes requires metadata (parameters) to control the flow of a process. Reuse of data products requires clear metadata to support its use within quality and privacy constraints. All three articles have some form of data and metadata management platform as a key component of their vision. These data service centers or common metadata databases ensure a consistent, reliable management of data and metadata with high quality. Statistics New Zealand highlights a key set of attributes for their platforms which enable the creation of statistical processes through specifications with little to no custom IT development required. This puts the power of process design in the hands of methodologists and statisticians who can create new capabilities in a flexible and timely manner.

Statistics Netherlands highlights the importance of managing "steady-states", and starting the process of design from the analysis of outputs, then designing the data, and finally the process. Chain management is highlighted to provide cross-process coordination. Data Service Centre concepts are introduced to provide careful management of these "steady states". Statistics New Zealand similarly notes the importance of data and metadata management, and highlights innovative approaches they have taken to associate powerful metadata with their data records. In addition, they highlight a shift from processing-centered data to analysis-centered data with the adoption of OLAP-based techniques optimized for analysis as opposed to more conventional OLTP-based techniques, which favor transaction throughput. They adopt powerful architectural principles (e.g., "Administrative data first"), reflecting their transformational shift away from traditional collection. NASS places the creation of centralized data and metadata management at the heart of their vision, remediating the traditional decentralized and *ad hoc* approaches used in their distributed field office environment. They leverage their infrastructure centralization and virtualization to create a core set of data and metadata databases, and similarly reflect on differences between OLTP and OLAP implementations.

Collectively, the focus on data and metadata management and use implies the need for a central information architecture, which provides the reference framework for governance, standardization, and specification to support these activities.

3.6. Processing

Several aspects of the survey processing are being improved by the three NSOs. Selective editing is one of these. This technique contributes to the reduction of the amount of manual work and follow-up activities devoted to observations that have negligible impact on statistical products. In the same vein, an iterative process – also referred to as responsive design – is proposed by Statistics New Zealand to derive estimates with a reduced set of sampled units. This approach uses quality indicators to determine whether collection activities should be modified (maybe interrupted) or not, in order to dedicate the limited resources available to where they will have a higher positive impact. Statistics Canada works on a similar approach based on rolling estimates (Ravindra 2012). Under this scenario, estimates will be tabulated regularly during the collection phase until they reach a desired level of quality, as measured by some indicators, at which point collection would stop. Along with a selective editing strategy, the scenario will contribute to reduce the number of manual interventions by analysts. Furthermore, it will highlight a connection between the collection process and the quality of the estimates.

3.7. Research Activities

The standard methods and systems that are essential to addressing the current challenges faced by NSOs need to continuously evolve in order to further improve efficiency or respond to future needs. To effectively do so requires research and development, as well as an architecture that can easily add new functionalities or their extensions into the integrated production lines. This enables an evolutionary state of the architecture after it goes through its transformational state. Regardless of any transformation initiative, a minimum support of research and development activities is essential, or the statistical organization will not survive the changes observed in the society it serves.

The role of methodologists is to steer statistical research and development priorities. Typically, the development cycle should go through four main phases: (1) idea generation, (2) proof of concept, (3) prototyping, and (4) integration. The methodologist leads the first three phases in order to contribute to specifications for the fourth phase.

On the other end, the IT analysts steer technical priorities. They are responsible for identifying technologies that can bring benefits to the production of statistical products. This research aspect enables IT development based on methodological specifications. A good symbiosis must exist between methodologists and IT analysts because technological and theoretical challenges are interdependent.

The strategy of the High-level Group for Strategic Developments in Business Architecture in Statistics (HLG-BAS 2012) is to promote international cooperation in which NSOs would share tools in order to reduce development efforts. Sharing can apply at multiple levels. Sharing of architectures and interface specifications can lead to more rapid development and enable future convergence. Co-development of components can lead to an open source approach allowing the community of interest to enhance elements in a controlled fashion. Sharing of executable modules is also effective at saving effort, although it is limited due to the fact that inevitably the owner is the only one who may evolve the module. In the end, the relative maturity of participating organizations and the catalogue of existing components each brings to the table will determine the right mix of approaches.

3.8. Evaluating the Transformation

Although most transformation projects are still ongoing, NSOs should already have thought about the way they plan to evaluate them. Lessons learned are documented but the success of the projects is not measured against initial objectives. An evaluation framework would highlight both the positive and negative outcomes of the transformation, such as: identifying key factors (methodological, technological or environmental), elements or decisions that contributed to successes where other similar initiatives had failed or been less successful in the past; better understanding of weaknesses in the planning or implementation phases; or, recognizing limitations in the transformation itself. This information is essential to improving the governance for future initiatives.

4. Concluding Remark

The three articles from NSOs describe transformations of survey functions, production lines, and management practices that are taking place or being considered in many NSOs. These transformations are essential to ensure that NSOs continue to deliver the information needed by the society that they serve for the years to come. These transformations are ambitious and some will take years to be fully implemented. In order to be sustained, they will have to deliver benefits along the way, even very early on. To achieve this, difficult decisions will likely have to be taken and some may even consist in prioritizing objectives. The key element is to set this transformation on an architecture that will allow the more standardized tools, methods and processes to grow and evolve through research and innovation in a well-governed manner. It is important to communicate from the onset that the new steady state will be globally better, that is to produce as much, if not more, relevant information of better quality at lower costs. It is also important to recognize that this information will be produced differently with fewer locally optimal features. Finally, some may resist standardization, integration and streamlining of processes claiming that it curbs research and innovation. On the contrary, such a transformation is innovation by itself and research is essential to achieving the objectives of the transformation and ensuring that the new steady state meets future challenges.

5. References

HLG-BAS (2012). Strategy to Implement the Vision of the High-level Group for Strategic Developments in Business Architecture in Statistics. Document presented at the Sixtieth plenary session of the Conference of European Statisticians, UN-ECE, June 6–8, 2012. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.58/2012/10_-_HLGBAS_Strategy.pdf (accessed March, 2013)

MIT-ESD Symposium Committee (2001). ESD Terms and Definitions – Version 12. Working paper ESD-WP-2002-01, Massachusetts Institute of Technology, Engineering Systems Division, October 19, 2001. Available at: http://esd.mit.edu/wps/esd-wp-2002-01.pdf (accessed March, 2013)

Ravindra, D. (2012). The Integrated Business Statistics Program: An Overview. Proceedings of the Fourth International Conference on Establishment Surveys. Montreal, June 11–14, 2012. Available at: http://amstat.org/meetings/ices/2012/proceedings/KESIV_TOC.pdf (accessed March, 2013)

Journal of Official Statistics, Vol. 29, No. 1, 2013, pp. 203–211

Discussion

Siu-Ming Tam[1] *and Bill Gross*[1]

1. Introduction

We congratulate the Editors for the production of this special JOS issue on "Systems and Architecture for High-Quality Statistics Production", and the selection of articles for inclusion in this volume. Statistics organisations, and national statistical organisations (NSOs) in particular, are at a crossroad. Faced with significant statistical challenges that have not been seen before (Pink et al. 2009), they can choose to transform their business model to "survive and thrive" or, ignoring these challenges, to continue to operate the way they have been successfully operating in the past and face the increasing risk of becoming irrelevant in the future.

In order to sustain the mission of NSOs to provide the statistical evidence for informed decision making, the choice is abundantly clear. Many NSOs have started thinking about or are embarking on a course with a view to offering new statistical products and services. They are finding creative ways to harness new data sources (a vision for products) and are actively looking for more efficient and effective ways in the production of statistics (a vision for processes). They also aim to maintain the same or a similar level of quality of the products.

Undertaking business transformation of the statistics production process and international collaboration in the development of concepts, classifications, tools and products are increasingly seen as promising strategies to address these challenges. Recognising the need to work collaboratively with NSOs in the world, the United Nations Economic Commission for Europe (UNECE) and, in late 2012, the United Nations Economic Commission for Asia and the Pacific (UNECAP) have set up High Level Groups comprising Chief Statisticians and Directors General from national and international statistical offices to work with member countries to develop an international vision and strategy for business transformation (HLG BAS 2011).

In addition, currently led by the Australian Bureau of Statistics (ABS), the NSOs from Canada, Italy, New Zealand, Norway, Sweden, and UK are members of a Statistical Network established to develop and experiment with models for working together to co-design and co-develop statistical frameworks, concepts, and tools. One of the more notable

[1] Methodology and Data Management Division, Australian Bureau of Statistics, Locked Bag 10, Belconnen, ACT 2617, Australia. Emails: Siu-Ming.Tam@abs.gov.au and Bill.Gross@abs.gov.au

Acknowledgments: The authors wish to thank Chris Duffy, Alistair Hamilton, Michael Glasson, Brian Studman, and the reviewers for their comments on earlier drafts of this article. The responsibility for the content of this article remains, however, solely with the authors. Disclaimer: the views expressed by the authors do not necessarily reflect those of the Australian Bureau of Statistics.

achievements of the Statistical Network is the development of the Generic Statistical Information Model (GSIM) (Hamilton and Tam 2012a). This is a metadata model for describing the information objects used in the statistical production process, as described by the Generic Statistical Business Production Model (GSBPM) (Vale 2009).

We will restrict our discussion to the Group of Four (GoF) articles by Nealon and Gleaton, Seyb et al., Struijs et al., and Thalji et al. The reason for our choice is principally a reflection of our interests in this challenging and emerging area in the work of NSOs. We will refer to these articles as the NASS (National Agricultural Statistics Service), SNZ (Statistics New Zealand), SN (Statistics Netherlands), and RTI (RTI International) articles respectively. Clearly the four articles cover a lot of ground in describing their journey to transform their business, and it will not be possible to discuss all aspects. Instead we chose to discuss the strategic issues in the business transformation, how they align with the emerging international vision for transformation, and conclude with comments on what we believe are significant challenges facing the NSOs as a statistical community.

2. The Products Vision – the GoF Articles

There are more similarities in the visions articulated in these articles than differences. The SNZ and SN articles describe efficiency, responsiveness, and expanding data sources as major aims for their business transformation. On the other hand, the NASS article articulates efficiency as the main aim, and achieves this through economy of scale by centralising their collection and network operations in a small number of offices in the NASS. The RTI article outlines improvement in accuracy and timeliness of their statistical products and services as their main aim, and seeks to achieve it by developing and implementing data collection and management systems to manage the production life cycle of statistics.

Whilst we generally agree with the articulated vision of transformation in the GoF articles, we feel that, as the articles are predominately focussed on the process vision, there is a gap in information about their statistical offerings. We argue that the products vision is the *raison d'etre* for the process vision and, without a vision for the product, the process vision can degenerate into a mere pursuit for efficiency. Whilst efficiency is important, we believe that innovation and creativity in harnessing new data sources to produce richer, more relevant statistics more responsively are the main challenges facing official statisticians now. The four emerging IT mega trends of Cloud, Social, Big Data, and Mobile provide major opportunities.

We therefore believe that readers of the GoF articles would benefit from learning their plans for the creation of new or better statistical offerings through the strategic use of the ever-increasing alternative sources of data: of administrative data available from governments, transactional data from businesses, and Big Data (or "organic data", Groves 2011) from social media, as well as through the opportunities available from the IT mega trends. That said, we felt that SNZ and SN are better positioned than NASS and RTI to do so, given the process transformation being undertaken in these organisations, a point we will return to later in this discussion. An excellent description of the characteristics of Big Data in terms of volume, velocity, variety, and degree of structure can be found in Daas (2012).

In articulating their products vision, the UNECE High Level Group stated:

> "The raw materials, the data we use to create our products, will need to be found in the data that are available throughout society. . .The active pursuit of data and the creation of products that give insight from an impartial perspective, our unique selling point, will be our new mission. It will undoubtedly mean that our organisations will have to leave their comfort zone and will have to question the work that seems so normal at present . . ." (HLG-BAS 2011)

There is a paucity of novel and innovative combinations of these different data sources, both existing and emerging ones, in the creation of new statistics. In the case of Big Data, Glasson (2012, internal ABS report) provides an interesting insight into why this may be the case by noting the differences in the supply chain and value proposition of these data sets between statistical agencies and nongovernmental organisations. There are some examples of a Big Data application for official statistics. Hal Varian, Chief Economist of Google, demonstrated in his lecture to ABS staff in 2011 how using Google search terms and statistical modelling can be used to "nowcast" US unemployment. The methods used in nowcasting are available in the public domain (Varian and Choi 2009).

Kimura et al. (2012) also describe a collaborative project between the National Data Centre of Japan and NTT DOCOMO, a Japanese major network operator, to use the latter's operational data about location and attribute to provide the capacity to estimate the daytime population on a real time basis.

Other examples of uses and issues with Big Data can be found in the Volume of Significance published in August 2012 by the Royal Statistical Society of the UK and in Daas (2012).

We urge statistical offices to release their plans for the creation of such new statistical products and services in the public domain with a view to sharing information and knowledge with their counterparts. In the ABS, our plans include the development and implementation of interdepartmental protocols for the creation of linked datasets combining ABS data with administrative data, subject to appropriate approvals, to provide the information to address complex issues facing contemporary Australia, and the development of facilities for accessing such datasets for secondary data analysis without compromising the confidentiality of respondents' data (Tam et al. 2009; Tam 2011).

3. The Process Vision

There is a wealth of information provided in the GoF articles on the transformation of statistical processes, ranging from organisation-wide transformational change to structural/systems change impacting on the collection and processing of information. We are particularly interested to compare their process vision with that of an emerging international vision for statistical transformation.

The process vision as articulated by UNECE High Level Group stated:

> "The production of statistics should be based on common and standardised processing. . .We view this as the industrialisation and standardisation of statistics production. . .The production of official statistical information should have its own industrial standards . . ."

Summarising these aspirations, the ABS recently used the slogan "Industrialisation, Modernisation, Transformation, and Future-Proofing" (IMTP) to galvanise support for its business transformation program across the ABS, and embarked on a strategy based on the following philosophies:

- plug and play – to build modular components (e.g., for imputation or weighting) which are easy to assemble into a reliable processing system, with an aim to supporting a future "assemble-to-order" statistical business model.
- standardisation and corporatisation – to standardise and corporatise concepts, classifications, methods, tools, and processes to reduce cost and lead time in the creation of statistical production processes and to improve timeliness and relevance of ABS statistical outputs. The effort to standardise and corporatise ABS metadata is huge and should not be underestimated.
- metadata-driven processes – to build content-agnostic processes whose metadata-rich outputs are driven entirely by input metadata.
- connectivity – to ensure statistical data and metadata can flow freely from one modular component of a statistical system to another with little, if any, manual intervention.
- capability building – to equip staff with the skills in anticipation of the move from low value-added to high value-added statistical activities; skills of which will be needed following the completion of the transformation, and
- international cooperation – to share or collaboratively build common standards, methods, tools and processes to decrease the ABS costs of their development (if going alone), and to work productively with the Statistical Network to achieve these aims.

Refer to Hamilton and Tam (2012b) for a more detailed exposition of these ideas.

3.1. GoF Transformation Strategies

Whilst all four statistical organisations plan to standardise tools, SNZ and SN aspire to build "plug and play" capability in their processes and tools with an aim of assembling components to form end-to-end statistical process systems quickly and with little or no IT involvement. In contrast, NASS and RTI have more modest goals and aim to "consolidate and generalize survey applications" (NASS) and for configuration of a particular collection to be "done by a programmer with less experience" (RTI). Building "plug and play" capability is a substantial capital investment and for statistical agencies with a large number of collections, there is ample opportunity to reap the budgetary benefits of such investment. Equally important, particularly for centralised statistical agencies, is that this plug and play capability enables new collections – whether direct, from administrative or transactional data or repurposing of existing holdings – to be developed and made operational quickly so that emerging user needs can be met in the time frames they expect.

While SN's (IT) treatment of metadata is unclear, the other organisations have information about data, with SNZ storing "metadata close to the data" but "in a way that facilitates data integration" and the others having a central store. SN and RTI extend the traditional thinking of metadata covering classification and data items to statistical processes, but NASS and SNZ handle information about processes and workflows separately. In addition, RTI has metadata about data, processes and workflows in a central system, and this has a number of advantages. It allows a process to update metadata about its outputs automatically and provides a comprehensive store from which to generate metadata for dissemination. It is easily searchable, thereby facilitating repurposing of data, reuse of workflows and use of standard processes. A central system should encourage metadata to be created in a standardised fashion.

We believe that metadata should be extended to cover paradata (Scheuren 2005) and metadata about statistical processes, as they are important to support an assemble-to-order business model as described above. Whilst we are not necessarily wedded to a centralised metadata store, federated data stores need to ensure that the metadata in all the stores are searchable and accessible by all statistical production staff. Furthermore, good metadata registration is needed for maximisation of metadata re-use.

Notably, SNZ makes extensive use of tools developed by other statistical agencies, such as BANFF (Statistics Canada) and Blaise (from SN), which is also used by NASS. The uptake in the use of tools developed by other NSOs is less than one would have liked to see. Part of the reason is that the tools that could be used in an organisation are constrained by the base level language (e.g., R for SN) or statistical package (e.g., SAS for Statistics Canada, SNZ, ABS) and associated data storage. More widespread use of these tools in the "future world" seems to require the developers to recognise the importance, and the provision, of "plug and play", "connectivity", and "metadata-driven" capabilities in the tools.

Summarising the lessons learnt in implementing a similar transformation program undertaken in the Office of National Statistics (UK), Penneck (2009) emphasised the importance of leading and sustaining a cultural change in the "new way of doing things". Whilst some references were made in the GoF articles about engagement with their business counterparts and statistical staff in general, there is not a lot of information, if any, in these articles to assist those who want to learn from their experience. We urge the authors make them available in the public domain.

3.2. Enterprise Architecture (EA)

A business transformation program is fundamentally a change program. To be successful in any change programs, it is important to articulate and communicate the vision and create a roadmap for change, including the creation of short-term wins (Kotter 1995). Successful change programs in an NSO will, in our view, require transformation of the business process, of how statistical data and metadata are managed, and of the technology infrastructure and application platforms. Outside the statistical world, it is customary to use an EA (Zachman 1987) as a tool to describe the interplay between, and steer, these essential elements of the change program. We believe that it is a useful tool for NSOs in undertaking their transformation programs.

From The Open Group's (2009) Architecture Framework, "[t]he purpose of enterprise architecture is to optimize across the enterprise the often fragmented legacy of processes (both manual and automated) into an integrated environment that is responsive to change and supportive of the delivery of the business strategy." It is customary to visualise the EA to comprise a Business Architecture layer, an Information Architecture layer, an Application Architecture layer, and a Technology layer, typically referred to as the Enterprise Domains. At the strategic level, the EA also comprises business segments including statistical leadership, statistical production, statistical support, and business support, which transcend people, products, and performance in the statistical organisation.

It is also customary to accompany the EA with a set of EA principles. These are used as a "guiding light" when it is unclear which one of a number of alternative choices, or pathways in the transformation, should be adopted. The SNZ article provided a set of the principles for their "Economic statistics business architecture", but we believe a broader set covering all official statistics would, for obvious reasons, be preferred.

4. Methodology and Methodology Architecture

Whilst all articles in the GoF mentioned statistical methods in conjunction with tools in the context of standardisation, the focus of the articles is on the tools, rather than the methodology. What role did methodology play in the transformational change carried out in the four statistical organisations, and how did methodologists contribute to the transformation process?

We believe that methodologists play a crucial role in shaping and contributing to the development and implementation of strategies to support a business transformation program. In recognition of this, we believe that the EA for statistical agencies should be extended to include a Methodology Architecture (MA) within the Business Architecture layer. Underpinned by the NSO transformational needs, the MA lays out the vision for methodology, supported by an envisaged future inventory of statistical methods and uses, contrasting with the current inventory, and a transition plan for migrating from the current methodological state to the future state.

In developing the MA, we believe consideration needs to be given to the following issues:

- how the suite of methods should be modularised – what should the scope and boundaries of each module be and what are the interdependencies?
- how to trade off between ease of application (usability), scope of application (flexibility), efficiency, and simplicity?
- the availability and use of process quality measures.
- the soundness and consistency of relevant conceptual frameworks underpinning the methods, and
- the relationship between MA and the Business, Information, Application and Technology layers within the EA.

5. Challenges for the NSOs as a Community

Pink et al. (2009) argued that given the enormity of the transformation program, it will be difficult, and indeed not sensible, for NSOs to do it alone. In the transformational space, we believe that NSOs could work (better) together in the development of the EA framework, standards (as was shown in the development of GSBPM and GSIM), products, ways to improve skills, and the processes to manage the transformation. In what follows, we outline what we believe are key challenges for the NSOs in collaborating as a community.

5.1. Collaboration in the Development of Future Statistical Products and Services

The use of Big Data as shown in Varian and Choi (2009) and Kimura et al. (2012) to create timely statistics relevant to official statisticians demonstrates the potential value of these data sets to contribute to the NSOs' mission to provide timely and relevant statistics and services for more informed decision making. One question faced by NSOs is: "Is this a space that we should be in?" In our view, the answer is obvious if we want to remain relevant in the future. If so, the question is how. Until recently, official statisticians gave little attention to Big Data's potential. However, this is going to be changed with the UNECE signalling its intention to commission a paper to understand what Big Data is and how it can be used in official statistics.

The statistical community will benefit from more sharing of ideas, developments and advances in this space, including the methods for processing and quality assuring the data.

5.2. Collaboration in the Development of Business Architecture (BA)

The GoF articles and our interactions with statistical counterparts in other parts of the world show that there is general convergence in our thinking on improving the efficiency and effectiveness of, and growing, our business. There is also remarkable similarity between the ideas about reusing data, reusing metadata, and reusing statistical methods and processes to improve the NSOs' value proposition to their users. Their vision is in strong accord with the vision of HLG-BAS.

Is the commonality of purposes for transformation in the NSOs of such a high level that it warrants the development of an international "reference" for statistical BA? If so, what work is needed for this to occur, who should be leading this work and how can NSOs contribute? The benefit of such an international collaboration effort is enormous, including the placement and recognition of statistical production standards like GSBPM, GSIM in this architecture. In late 2012, UNECE released GSIM 1.0 (UNECE 2012), including a User Guide and specification for consultation within the international statistical community. The Statistical Network referred to earlier has commenced a project to collaborate on developing a "common" BA for the participating countries in the Network.

5.3. New Form of International Collaboration to Create Statistical Processes and Tools?

Hamilton and Tam (2012b) argued that official statisticians can learn a lot from the model used by some well-known smartphone manufacturers in getting the world's developers to develop applications for their handsets to improve the functionality of the phones and user experience. They argued that this collaborative model should be added to the "stand-alone" model (as used in the development of PCAXIS, Blaise, etc.), the "co-development and co-design" model (as used in developing GSIM), and other existing models for international collaboration.

Under this "platform-based" collaboration model, developers (e.g., staff in the NSOs) can upload applications to an internet-based statistical applications store for sharing. In choosing the processes or tools to plug and play, the official statistician shall no longer be restricting the search to applications within their own statistical organisation, but will have applications from other NSOs available to search and to download. In addition, private developers who want to sell their products to NSOs can also upload them to the same store for showcasing to the NSOs.

If this new form of international collaboration is to work, a number of prerequisites have to occur, including a reference EA, with internationally agreed protocols for developing applications under the EA, standards for plug and play, connectivity, and metadata-driven systems. Whilst some work has been undertaken on this front, more needs to be done and particularly more involving the great many NSOs outside Europe, Asia, and Pacific, and beyond the participating countries in the Statistical Network.

In regard to defining standards for "plug and play", we are pleased to note that, in November 2012, HLG BAS commissioned a major collaborative work program to be undertaken in 2013.

We thank the authors for their thought-provoking articles.

6. References

Daas, P. (2012). Big Data and Official Statistics. Sharing Advisory Board Newsletter, Issue 7. 2–3. Available at: http://www1.unece.org/stat/platform/download/attachments/22478904/issue+7.pdf?version=1&modificationDate=1354288238441 (accessed February 12, 2013).

Groves, R. (2011). Three Eras of Survey Research. Public Opinion Quarterly, 75, 861–871.

Hamilton, A. and Tam, S.-M. (2012a). Toward GSIM V1.0 as a Cornerstone for Common Reference Architecture. Meeting on the Management of Statistical Information Systems (MSIS 2012) Washington, DC, 21–23 May 2012. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2012/02_Australia.pdf (accessed February 12, 2013).

Hamilton, A. and Tam, S.-M. (2012b). Platform for International Collaboration – Preliminary Thoughts from the ABS. Meeting on the Management of Statistical Information Systems (MSIS 2012) Washington, DC, 21–23 May 2012. Available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2012/20_Australia.pdf (accessed February 12, 2013).

HLG-BAS (2011). Strategic Vision of the High-Level Group for Strategic Developments in Business Architecture in Statistics. UNECE. Available at: http://www1.unece.org/stat/platform/display/hlgbas/Strategic+vision+of+the+High-level+group+for+strategic+developments+in+business+architecture+in+statistics (accessed February 12, 2013).

Kimura, M., Makita, N., Terada, M., and Kobay, M. (2012). Developing Value-Added Small Area Statistics with Operational Data of Mobile Network. Paper presented at the Conference of the International Association of Official Statistics, Kiev, 12–14 September 2012.

Kotter, J. (1995). Leading Change: Why Transformation Efforts Fail. Harvard Business Review, 73, 59–67.

Penneck, S. (2009). The Office for National Statistics Statistical Modernisation Programme: What Went Right? What Went Wrong? In Proceedings of Modernisation of Statistics Production, International conference organised by Statistics Sweden, Stockholm 2–4 November 2009.

Pink, B., Borowik, J., and Lee, G. (2009). The Case for an International Statistical Innovation Program – Transforming National and International Statistics Systems. Statistical Journal of the International Association for Official Statistics, 26, 125–133.

Scheuren, F. (2005). Paradata from Concept to Completion. In Proceedings of the Statistics Canada Symposium 2005. Ottawa: Statistics Canada.

Tam, S., Farley-Lamour, K., and Gare, M. (2009). Supporting Research and Protecting Confidentiality – ABS Micro Data Challenges and Future Directions. Statistical Journal of the International Association of Official Statistics, 26, 65–74.

Tam, S. (2011). On-line Access of Micro-data at the Australian Bureau of Statistics – Challenges and Future Directions. Proceedings of the 58th World Statistics Congress of the International Statistical Institute.

The Open Group (2009). TOGAF Version 9. Available at: http://www.opengroup.org/togaf/ (accessed February 12, 2013).

UNECE (2012). GSIM Version 1.0. Available at: http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=59703371 (accessed February 12, 2013).

Vale, S. (2009). The Generic Statistical Business Process Model, Version 4.0. Available at: http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model (accessed February 12, 2013).

Varian, H R. and Choi, H. (2009). Predicting the Present with Google Trends. Google Research Blog. Available at: http://googleresearch.blogspot.com/2009/04/predicting-present-with-google-trends.html (accessed February 12, 2013).

Zachman, J. (1987). A Framework for Information System Architecture. IBM System Journal, 26, 276–292.

Journal of Official Statistics, Vol. 29, No. 1, 2013, pp. 213–220

Discussion

Dennis Trewin[1]

1. Introduction

To provide context for these remarks, I was Chief Executive of the Australian Bureau of Statistics from 2000 until 2007. During this time we introduced major changes to our systems for the collection, processing and dissemination of business statistics involving significant re-engineering. It also involved major changes to our organisational structure. Our inspiration came partly from discussions with Statistics Netherlands in 2003 and we started on the strategic design work shortly afterwards. This project was extremely successful – the objectives were achieved and it was completed on time and on budget. I wanted to outline the reasons I believe this was the case because I think it is relevant to the discussion of the articles in this special issue. In particular, I use our positive experience as something of a benchmark in analysing the production systems in this special issue.

First, the desired *objectives* of the redesign were quite clear and mutually supportive. They were:

- improved data quality.
- improved relations with businesses, improved reporting mechanisms and reduced reporting load.
- facilitating the increased use of administrative data, particularly taxation data.
- increased capacity to respond to emerging statistical demands.
- significant operating efficiencies by reducing the number of staff by 30 per year (about 15% per year), and
- enhanced opportunities for staff by creating units that focus on statistical outputs not statistical collections.

We were greatly assisted by the fact that we had received a significant increase in access to taxation data as a result of the introduction of a Goods and Services Tax (GST). This new tax was supported by a quarterly Business Activity Statement (BAS) for most businesses (monthly for the largest businesses). The number of items was relatively small but it contained the most important data items for supporting the quarterly national accounts. In effect, we had a quarterly census of businesses.

Second, there was extensive *consultation* with the different layers of management and the staff who were to be most affected by the changes. We started by having a small team

[1] Statistical Consulting, Address: 38 Mirning Crescent, Aranda, ACT, 2614, Australia. Email: dennistrewin@grapevine.net.au

developing strategic scenarios to achieve these outcomes for discussion by management. A preferred strategy was developed and discussed more generally with staff. These discussions were led by me and my most senior colleagues. Further changes were made as a result of those discussions. The fact that we made changes surprised staff somewhat and helped to develop strong buy-in from them and the staff unions. The importance of these consultations proved to be much greater than we anticipated. The union subsequently referred to our approach when discussing development projects with other government agencies.

Third, one of the biggest concerns expressed by staff during consultation was the need for *reskilling*. As a consequence, early attention was given to identifying the reskilling needs and an implementation strategy.

Fourth, the project had the very clear *support of executive management*. There were a small number of doubters and some effort was put into understanding the reasons for the doubts and what could be done to address them. In the end, there was 100% support for what was a major redevelopment project.

Fifth, early attention was given to *project management* (Trewin 2007). An external expert was used to provide advisory services throughout the project. A project management framework had been established and was used for this project. In particular, a lot of attention was given to the project governance arrangements. A project office was established to support the multiple project management committees and project teams.

Sixth, it was clear that significant *re-engineering* was required if the project objectives were to be achieved. Early attention was given to the nature of the changes so that organisational changes, reskilling requirements, and so on were understood. In particular, an early decision was made on the restructuring that was to be undertaken. This enabled us to take special steps to maintain business continuity whilst migrating to the new business structure. We looked for a mechanism that would ensure, as far as possible, that functions to be detached from existing collection areas were well defined, well understood and able to be managed as a distinct entity prior to transfer. This was done by creating "pods" working in the new way but located within the existing business units before the migration took place.

Seventh, a special feature of the project management arrangements was to identify the key *risks* and how they were managed. This was done for the overall program and individual projects. Possible business discontinuity was identified as the key risk, especially by the statistical managers. Special steps were taken to avoid this, including the podding described in the previous paragraph.

Eighth, *engagement with clients* of any new production system is crucial. In this case these were the managers and staff responsible for the new system. A lot of effort was put into consulting with these stakeholders prior to development and during the development period.

Ninth, there was a strong *focus on outcomes*, not just outputs. For example, we had past experience where systems had been developed in accordance with specifications but, for various reasons, could not be used by those who were intended to use them or did not perform effectively. That is, outputs were delivered but the desired outcomes of the systems redevelopment were not achieved. This particularly came into play when considering the training requirements, acceptance testing, documentation, and so forth to ensure the production systems were fit for use and could be used before they were formally installed.

Finally, we stayed with *proven technology* even though it was relatively new from the point of official statistical agencies. One of the core technologies used was input data warehousing. The ABS' enterprise architecture was updated as a result of decisions on technology to be used for the new production systems.

2. The Production System Proposals

I will now look at the Statistics Netherlands (SN), RTI International (RTI), National Agriculture Statistical Service (NASS) and the Statistics New Zealand (SNZ) proposals for each of these ten criteria. The Westat/National Centre for Health Statistics (WESTAT/NCHS) proposal is considered separately because it is concerned with data access rather than data collection, processing and dissemination.

With respect to the *objectives* there was a strong focus on cost savings in all the proposals. This is not surprising, but I wonder if it means the other important objectives, especially over the longer term, are not getting sufficient attention in some of the proposals. In particular, is there enough focus on the dimensions of quality (e.g., accuracy, timeliness, coherence)? There are some exceptions. The SNZ proposal does provide a focus on a broad range of specified objectives, including supporting analysis capacity for SNZ statisticians, not just traditional processing functions. The SN proposal puts emphasis on flexibility to meet emerging user needs. The RTI and NASS proposals put a lot of emphasis on paradata as a means of improving the effectiveness of their survey operations.

Standardisation and integration were common objectives of all the proposals as a means of reducing cost, but this was being implemented with varying degrees of success. A range of different forms of standardisation are proposed – methods, concepts (classifications, definitions, etc.), processes, and supporting tools and systems. All are important. Standardisation of methods and concepts is important to the statistical integration increasingly desired by the users of these statistics, including the national accounts. The use of standard tools and systems, dependent to a large extent on standardisation of processes, is an important part of standardisation but it can be difficult to persuade some to move from a familiar system they have been using to a corporate standard, even if there is no loss in functionality.

With respect to *consultation*, employee ownership and involvement was only specifically discussed in the NASS proposal. It may have been considered in the other proposals but its non-mention suggests it may not be getting sufficient attention.

Reskilling is crucial to the success of production systems that involve major changes and got a mention in most of the proposals. For example, the SNZ proposal said that they would need to address the changing skills required by subject matter statisticians and the RTI proposal highlights the need to develop skills in using paradata effectively. A major change to production systems is more likely to be successful if the reskilling needs are identified early and plans to address these needs are made.

The clear *support of executive management* is essential for the success of major development projects. It can be difficult as their successful implementation often requires cultural change. The corporate benefits have to be understood even if some "local" inefficiencies are caused by using corporate systems in the short term. There will be resistance, but it may take significant effort to understand the reasons for the resistance and what needs to be done to convert the resisters. Often, it will be necessary to change personnel or their responsibilities, but this should be considered as a last resort.

The necessary support of executive management has not been obtained in all the production system proposals. The SNZ article recognises the need for corporate decisions on statistical development projects and the importance of cultural change but, at the time the articles were written, both the NASS and RTI proposals suggest there is some resistance that may not yet be resolved. In the NASS case, it says the Review Survey Estimates application has been delayed because maintenance of legacy systems has taken priority. In the RTI case, the statement that "Although RTI senior management has the expectation that each new survey project will adopt and implement Nirvana, the extent to which this happens – and reasons for doing so or not doing so – will be monitored" suggests the use of Nirvana will be non-mandatory, thereby reducing some of the corporate benefits.

The SN article says they have "not solved the standardisation challenge yet". They have been considering this type of approach to their production systems for some time (an ICT Master Plan was developed in 2005) and I believe the main reason they have not progressed further, or achieved the necessary standardisation, was that there was not the necessary executive support for the proposals. This may be for good reasons; perhaps the proposal should be modified in some way to obtain the necessary support? There have been some recent positive signs. The authors say: "when the notion of architecture was first introduced, it was met with scepticism and its added value was questioned". However, the authors now note that this scepticism has largely disappeared.

Project management, including project governance, is crucial but did not get detailed attention in some of the production system proposals. There is not sufficient space to provide a critique on each of the proposals, but the NASS article provides an excellent case study of good project management. Project management also got a lot of attention in the SNZ article. It is the SN proposal that concerns me most. Although multi-disciplinary teams will be used, the governance arrangements suggest there is a risk it may become an IT project rather than a statistical project enabled by IT. Certainly, using the CIO as Chair of the Central Portfolio Board rather than the "owner" of the production system being re-developed is unusual and is likely to lead to lack of ownership by the statistical areas.

However, some parts of the SN approach are worth considering in the design of major development projects. They used three different levels of project management which I support although, as mentioned above, I wonder whether the CIO should chair the Central Portfolio Board:

– managing the portfolio of redesign projects (chaired by the CIO).
– managing individual redesign projects (headed by statistical managers), and
– managing the enterprise architecture (chaired by the CIO but an Advisory Board to the Director-General).

The development projects being considered in this special issue are rather large. To make them more manageable, it is important to break them up into sub-projects, each with their own deliverables, milestones and budgets. This also makes it much easier to monitor progress and provides a sense of achievement when intermediate milestones have been reached. This can be important for large projects with rather extended timetables.

Re-engineering and restructuring will be crucial to obtaining the full benefits of the revised production systems, but are the staffing implications receiving enough attention?

Re-engineering can be difficult because it usually involves significant changes to staff responsibilities, but it is usually necessary to obtain the benefits of the new production system. Some re-engineering proposals are mentioned, such as the changes needed to effectively use paradata and introduce adaptive survey design (NASS and RTI); the need for greater integration of production systems (all proposals); easy to use standards, standard tools, and lower response burden through increased use of administrative data (SN and SNZ); and the need to move away from stovepipes and increase the use of administrative data (SNZ and SN).

Risk management is important to the success of any development project. Business discontinuity can be the most important risk with major redevelopments of production systems. Certainly, this was the risk that got most attention with the ABS' business statistics project. However, it did not receive much attention in any of the articles, although the NASS article recognised the importance of managing the transition to support business continuity and the RTI proposed a live field testing strategy. I have doubts about whether the bottom-up incremental approach proposed by SN to reduce risks is likely to be effective, as it will be difficult to maintain full corporate commitment to the redeveloped production system. This should not be taken as a criticism of pilot testing of the approach as an initial step. However, once the approach is proven the implementation strategy should encompass the transition of all the relevant production systems.

The clients for these four systems are the potential users of the system. They are the target audience when we talk about *engagement with clients*. It requires continual communication from the beginning to the end of the development. They need to be involved in the design of acceptance testing. The system cannot be considered complete until it has been signed off by the end users. These issues do not receive a lot of attention in the articles.

It is important to take account of the needs of the external users of the statistics derived from these production systems. However, this should not generally require a separate "user engagement" activity as their requirements should be known. The external users will require flexibility in the data inputs (e.g., inputs derived from questionnaires) to meet new and emerging needs. They will also have specific requirements in the way the outputs are delivered. The data input and dissemination components of production system have to be designed to allow this flexibility.

Similarly, a *focus on outcomes* rather than outputs did not receive much attention in the articles, although this does not mean this aspect was ignored. One exception was the NASS article, where the fact that "payment for services was not made to a contractor until the project manager was satisfied that the application was ready for production" clearly indicates an outcome focus.

Most of the articles suggest that *proven technology* will be used as the core technology for the production systems. This is explicitly mentioned in the NASS article. The SNZ article also suggests they will be making extensive use of off-the-shelf software. RTI carefully considered this approach but decided to develop their own systems without the use of off-the-shelf software. Even if proven technology is used, IT has to be managed to ensure standardisation occurs and is maintained. Indeed, as an example NASS mentions that issues have arisen because IT has not been well managed in the past. In fact, this might be a reason that some of changes they are making (centralised web servers) have not been used previously, even though this technology has been used for well over a decade by my former organisation (with eight regional offices). Standardisation offers the potential for considerable savings in IT expenditures. The SN article explicitly discusses how they plan to introduce an enterprise architecture to enforce standardisation. No doubt an enterprise architecture is behind the other proposals although not explicitly stated as such.

3. Some General Questions

Discussants were provided with some questions that they might address as part of their comments. The answers to those questions which are relevant to my comments are set out below.

In your view, what are the biggest breakthroughs, or achievements, that have been made in redesign of the systems, as presented in the articles? Standardisation is a common motivation for all the redevelopment projects. As stated above, a range of different forms of standardisation are proposed – methods, concepts (classifications, definitions, etc.), processes, and supporting tools and systems. All are important. Although standardisation makes great common sense, it is not always easy to achieve as there is often considerable resistance because of the reluctance to move from current familiar methodologies or systems which work even if out of date. However, the corporate benefits from standardisation are considerable. These include reduced operational costs including training costs, reduced IT expenditure, greater consistency of approach across collections leading to improved statistical coherence, and greater flexibility in moving staff from one statistical product to another.

In your view, what are a few of the biggest mistakes, oversights, or shortcomings that system designers make in redesigning their systems? Without doubt, poor project governance is the most common cause of project failure even when the concept is great. Insufficient attention to risk management is often an important aspect of poor project governance. It is reassuring that for most of the production systems in this special issue, attention appears to have been given to project governance. The SNZ article provides some analysis of the lessons learnt from previous attempts to introduce a standardised approach. Two things mentioned were the inadequacy of the budget and trying to undertake the task in too short a time frame.

What role should survey methodologists play in the redesign of system architectures? Methodologists have very important roles, as I describe in the following sentences, but they should not have the project leadership role. That is the responsibility of the project owners, that is, those who will be using the new production systems in their day to day work. The methodologists may be the source of the innovation behind the production systems. They bring a broad perspective and can see the common elements across data collections more easily than those working on specific data collections. It is also part of their responsibility to keep abreast of developments that are happening elsewhere, including other countries. For these reasons, they should play an invaluable role in the exploratory and development phases (including risk analysis). Statistical methods can also play an important role in the design of survey modules and should be involved in this work. They should be represented on the Project Boards and Steering Committees as they can provide a different perspective to most of the others on these bodies.

Management (or governance) of development projects is crucial but it is not a well documented or understood practice. Governance of large-scale development projects needs to take into account the idiosyncrasies of the organization, but there may be important elements that are common to many organizations. Do the articles provide any evidence of these common elements and can you elaborate on what they are as well as their importance? I do not agree that management of development projects is not a well documented practice. There is an extensive literature (see Lock 2007; Program Management Institute 2008; Williams 2008), although perhaps not for statistical development projects. However, many National Statistical Offices have dedicated training programs on project management. For example, the ABS has its own Project Management Framework based on standard project management approaches with complimentary training programs (see Trewin 2007). Perhaps it is more accurate to say that the special application of project management to statistical development projects is not well publicised in the statistical literature.

There is not a consistent interpretation of project management and project governance in the articles. In my remarks I have taken governance to be a (very important) component of project management. I believe this is consistent with the majority of the literature on project management.

4. The WESTAT/NCHS Article

I have commented separately on the article by Krenzke et al. because it deals with a completely different part of the production system, that is, external online access to microdata in a safe environment. Whilst this system is incredibly sophisticated, with very strong theoretical underpinnings, I have my concerns about whether it will be a successful development project. In my mind, it seems more complicated than necessary. The development cost will be extremely high and I wonder if it will be justified by the use. In fact, I think many of the users will be frustrated by the limitations on use:

– It is not clear how much consultation there has been with potential users of the system. Based on my experience in Australia, they will be disappointed with the limited number of functions that will be available to them.
– A risk avoidance rather than a risk management approach has been used. As a consequence, the system will be extremely complex and difficult to develop and maintain.
– The approach assumes users cannot be trusted. Should there be more reliance on legally enforceable undertakings? I am not clear what is allowed by the legal arrangements under which the National Centre of Health Statistics operates but, if there are provisions, it may be possible to simplify the requirements.
– Do users really give a high priority to accurate estimates of variance? I would have thought they would be more interested in increasing the analytical functions they can apply to the microdata.
– Remote access systems to microdata have been commonly used in many parts of the world for some time. There is no evidence that the strengths and weaknesses of these systems have been analysed and used to assist in the design of the WESTAT/NCHS system.

– Unlike the other production systems in this issue, little use of off-the-shelf software is proposed.
– The project governance arrangements are unclear. Without strong management, it would be very easy to lose control of such a complex project.

5. The Future of Statistical Production Systems

I believe the future will see a different form of cooperation between National Statistical Offices in the development of tools and methods for systems development. There has already been a lot of cooperation on methods through statistical conferences and/or bilateral and multilateral meetings. There has also been sharing of processing tools for production systems. The BLAISE system developed by Statistics Netherlands is a notable example.

The work taking place on process modeling and information modeling with the Generic Statistical Business Process Model (GSBPM) and Generic Statistical Information Model (GSIM), under the auspices of the Conference of European Statisticians, may underpin the future of Statistical Production Systems. It will support the application of common methods across statistical offices. It will support the development of common tools based on this model. This can be particularly important for developing counties. The licensing costs of commercial statistical software packages can be out of the range of many developing countries. The availability of tools at the relatively low cost based on a relevant GSBPM/GSIM will be a real bonus to them. However, there are too many idiosyncrasies in the requirements of individual countries to expect this work to lead to common systems.

These tools are likely to facilitate and depend upon the life cycle management of data and metadata. They are likely to be part of an agreed enterprise architecture. This will only work if there are effective governance arrangements for IT and statistical information management.

Finally, I would like to congratulate all the authors on excellent contributions to this very important topic for statistical agencies.

6. References

Lock, D. (2007). Project Management (Ninth Edition). Farnham: Gower Publishing Limited.

Programme Management Institute, United Kingdom (2008). A Guide to the Program Management Body of Knowledge. Programme Management Institute.

Trewin, D. (2007). A Project Management Framework that has Worked for the ABS. In Organisational Change and Project Management, John Wanna (Ed.), Canberra: Australia National University Press. Available at: http://epress.anu.edu.au?p=118751

Vale, S. (2009). Generic Statistical Business Process Model, Version 4.0. Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS), Eurostat, Luxembourg.

Williams, M. (2008). The Principles of Project Management. Melbourne: SitePoint.

Journal of Official Statistics, Vol. 29, No. 1, 2013, p. 221

In Other Journals

SURVEY METHODOLOGY

A Journal Published by Statistics Canada
Volume 38, Number 2, December 2012

CONTENTS