Journal of Official Statistics, Vol. 29, No. 1, 2013, pp. 157–163, DOI: 10.2478/jos-2013-0009

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/.