Distr.
GENERAL

WP.12
15 April 2011

ENGLISH ONLY

**UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE (UNECE)
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE EUROPEAN UNION (EUROSTAT)**

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD)STATISTICS DIRECTORATE

Meeting on the Management of Statistical Information Systems (MSIS 2011)
(Luxembourg, 23-25 May 2011)

Topic (ii): From local to corporate perspective (industrialization and standardization)

The Challenges of Consolidating Federal Government Information Systems and the Genesis of SCOPE[1]

Invited Paper

Prepared by Ronald F. Bianchi, Economic Research Service, US Department of Agriculture (United States)

I. Introduction

1. With the advent of the internet, its search tools, data mining techniques, powerful computers for merging data, and now the wide availability of numerous person-based data files, the U.S. (Federal) government is facing a large challenge in protecting data from disclosure and privacy breaches. Federal government agencies at all levels are reviewing regulations, policies and procedures due to the heightened level of disclosure risk in the current environment. Many sectors of the Federal government are actively promoting the release of even more government information in the spirit of transparency and openness. This effort is reflective of the administration's Data.gov website and the President's Open Government Directive. The balance between protecting data while releasing more data presents a difficult and ongoing negotiation that becomes even more complicated when Federal departments consolidate data and many data

[1] Statistical Community of Practice and Engagement

management functions. This risks commingling of data in information systems that could produce internal agency privacy breaches.

Privacy Security Consolidation.

Data Integrity Open Government Transparency Functionality

2. While it may be difficult to include all these terms in a single sentence, Federal managers in the realm of statistical data must continuously reconcile these multiple objectives on a daily basis. The President's Open Government Directive introduces a dynamic which runs headlong into the ongoing buttressing of the security of data residing on Federal systems. It grows ever more difficult to assure survey respondents that we can maintain privacy of collected data while being directed to be more transparent. Consolidation of data center resources within departments, while offering enhancements to overall system security, also enhances the probability of internal breaches resulting from unintentional technical acts. The Statistical Community of Practice and Engagement (SCOPE), sanctioned by the U.S. Interagency Council for Statistical Policy (ICSP) within the Office of Management and Budget (OMB, Executive Office of the President), was formed to address these and other issues confronting US statistical agencies through collaboration and sharing of best practices.

II. In the Beginning

3. Over six years ago I became the Chief Information Officer (CIO) for the USDA Economic Research Service (ERS). While attending ICSP meetings for my Administrator, I couldn't help but notice that issues existed between statisticians and information technology (IT) management within their parent departments. For example, boutique statistical software packages could take months to purchase as they were not on the standardized departmental IT list of acceptable software. Some agencies were strongly resisting attempts of their parent departments to consolidate their IT infrastructure and statistical data within the parent department's enterprise data center (EDC). Such consolidation has been mandated by the Federal CIO to attain efficiencies and economies of scale.

4. I saw this problem as an opportunity to convert the intra-department dynamic from adversarial to collaborative. By melding the interests of both IT and statistical management across statistical agencies, I hoped for a shift in their perspective to enhance the probability of success within both spheres of influence. Thus was born the Statistical Community of Practice and Engagement (SCOPE).[2]

5. An existing committee of IT and statistical staff, the Privacy Committee of the Federal Committee on Statistical Methodology, seemed the appropriate forum for planting the SCOPE seed, where it was enthusiastically embraced. We established a task force to investigate how such a group might operate and begin to brainstorm possible projects. We invited OMB E-

[2] Initially named Statistical Line of Business, to accommodate the prior Administration's E-Government approach, it was decided that a change was needed, if for no other reason than the acronym—SLOB—seemed a bit unfortunate.

government management to attend, as a major objective for the group was to level the playing field between advocates for statistical agencies' and e-gov's objectives. OMB provided and SCOPE adopted the same structure and documentary support used by other Federal government-wide lines of business, e.g., budget, personnel, payroll, records management. SCOPE task force members drafted the foundational document, the Concept of Operations, over the following months, and approved by the ICSP. The Director of OMB inserted $2MM into the President's FY 2011 budget to support a Program Management Office within ERS. Unfortunately, this funding has evaporated during the ongoing budget battles.

III. The SCOPE Vision, Goals and Objectives

Provide a collaborative community for statistical agencies to produce relevant, accurate, timely, cost-effective data and insightful research disseminated through shared state-of-the-art best practices to support data-driven decisions throughout the Nation.

The goals and objectives supporting the vision are:

Goal 1: Improve the availability of relevant, accurate, and timely Federal statistical data to all data users.

Goal 2: Improve the integration of data collection across Federal statistical agencies to increase interoperability and to minimize duplication of efforts.

Goal 3: Increase the use of data sharing across Federal statistical agencies to reduce respondent burden and to enhance the utility of existing government information systems.

Goal 4: Increase the capacity and improve capabilities for the analysis and research of Federal statistical data to support program decision making and policy development.

Goal 5: Enhance data security and access by aligning information technology support across Federal statistical agencies to protect the integrity of confidential Federal statistical data.

6. Briefly, the SCOPE process entails a workgroup, representing the principal Federal statistical agencies, agreeing on projects and budget for the upcoming fiscal year and submiting it to ICSP for approval. Approved projects would be led by statistical and IT staff volunteered by

member statistical agencies. Projects would be funded by agencies' contributions via a memorandum of understanding, direct appropriation, fee for service (e.g., use of software), and/or possible public/private partnerships (free hosting), and managed by the Program Management Office (PMO). Critical to meeting this vision is the protection of confidential statistical data. Allow me to use my agency as an example.

7. In our role as a federal statistical agency, data are critical corporate assets. As such, we must provide for data integrity and secure access to those who produce and use the data. ERS has a variety of research and database support services including mapping, data visualization, spatial analysis and modeling, data acquisition, data integration, and database development and management. Analysts often participate in multiple research groups providing consultation on data availability and spatial analysis techniques. ERS routinely uses this critical capability to deliver a range of high-visibility, relevant tools for use in the policy community and beyond. This capability has been used in modeling food deserts, estimating impacts of planting flexibility, and estimating the value of rural broadband. Most recently, this capability was used to design and develop the Food Environments Atlas as requested by departmental leadership and designed in consultation with White house staff to support First Lady Michelle Obama's efforts to fight childhood obesity. Given the mission-critical nature of such research and data activities, it is essential for ERS to retain logical control over servers containing the data that enables these successes.

8. SCOPE has been conceived to fill a gap by providing a government-wide capability to provide a more uniform, systematic, and coordinated means of communicating among the Federal statistical community. It will foster the sharing of information for the purpose of improving the statistical protocols and tools that support the production, analysis, and dissemination of statistical information. The mutual benefits derived from such inter-agency collaboration and cooperation will serve to benefit the data user community, while sustaining the protection of confidential data. A statistical community of practice empowered by a viable and sustainable operational capability in the form of the SCOPE initiative would provide an integrated means of defining opportunities and approaches and disseminating practical, effective solutions.

9. The majority of common cross-agency activities associated with data management and dissemination involve information technology infrastructure and architecture. For full collaboration across agencies, ready access to unique statistical software, shared data and software developments for those data would be a prerequisite. The recognition of the unique information technology needs of the statistical community, especially the ability to consolidate software acquisitions and to manage direct access to statistical data through a virtual data center, would be essential to the success of SCOPE.

10. The data collected, analyzed and disseminated by Federal statistical agencies are used by:

- policy makers at all levels of government to inform legislative and regulatory decisions,
- other government agencies within and outside of the Federal statistical community to supplement their data portfolios and to put the information they release in perspective,
- researchers to improve the knowledge base, and
- the general public to inform everyday decisions.

11. Efforts on the part of SCOPE to develop and leverage improved dissemination methods, formats, and presentation techniques that are cost-effective, and serve to increase compatibility and interoperability across agencies will facilitate user access to federal statistical information. By connecting collaborative work teams from federal statistical agencies, IT specialists, statisticians, and researchers could come together to advance this mission. Emerging technology will make it possible for these teams to collaborate more efficiently and effectively on a broader range of projects than has been the case in the past.

12. OMB's Office of Information and Regulatory Affairs (OIRA) provides a coordinating forum for federal statistical agencies through the Interagency Council on Statistical Policy (ICSP), of which ERS and the other federal statistical agencies are members. Under the auspices of the ICSP, ERS has been designated with the leadership role for SCOPE. As an initial step, ERS will serve as the Project Management Office for SCOPE. Modeled after OMB's Budget Formulation and Execution Line of Business, the SCOPE mission includes projects designed to ***enhance the horizontal, functionally-based integration of IT resources among Federal statistical agencies***. A statistical enterprise data center would be mandated to foster creative approaches for collecting, storing, analyzing and otherwise processing Federal statistical data to meet statistical agencies' missions—with significant cost savings—and to more efficiently feed data to Data.gov. The center would house Federal and commercial statistical datasets; visualization and dissemination tools; data quality, interoperability and confidentiality tools; and statistical analytical applications and models for cloud-type access by all Federal statistical agencies. This would also more quickly meet the Federal CIO's broader strategic objective of data sharing across Federal departments.

13. In other words, it is better to consolidate IT resources on a functional basis than arbitrarily consolidating vertically within parent departments. For example, under the vertical model, ERS data and IT resources would be consolidated with resources of the US Forest Service—very different missions, very different data and applications.

14. Most countries have a single statistical agency, achieving economies of scale. SCOPE offers a collaborative approach within a distributed system, attaining efficiencies and economies of scale without severing agencies' symbiotic subject matter-based relationships with their parent departments.

15. This rationale was proposed to ERS' parent department in an attempt to maintain control of its IT resources.

IV. It's the Law…

16. Most Federal statistical agencies have their own enabling statutes, within which are mandatory procedures for protecting confidential statistical data. But one statute prescribes the minimal parameters of protection for such data, **Confidential Information Protection and Statistical Efficiency Act.**

17. Title V of the E-Government Act, the Confidential Information Protection and Statistical Efficiency Act (CIPSEA), requires statistical agencies to maintain the confidentiality of information collected for statistical purposes under a pledge of confidentiality. In pertinent part, Section 512 (B) states:

> (b) DISCLOSURE OF STATISTICAL DATA OR INFORMATION -- Data or information acquired by an agency under a pledge of confidentiality for exclusively statistical purposes shall not be disclosed by an agency in identifiable form, for any use other than ***an exclusively statistical purpose***, except with the informed consent of the respondent.

17. A statistical agency would arguably be in violation of this statute to the extent it disclosed statistical data by transferring it to an EDC, as such a disclosure would be for a use other than an exclusively statistical purpose.

V. …As Well as Principles and Practices

18. According to the *Principles and Practices for a Federal Statistical Agency*[3], characteristics related to independence for a federal statistical agency include that the statistical agency have authority for the selection and promotion of professional, technical, and operational staff to carry out their statistical mission. If a statistical agency were to surrender access control to the IT organization of its host department, it would no longer control who can access its data or how those data are used. Thus, the agency would not only be violating its past pledges of confidentiality to respondents, but it would also no longer be able to make confidentiality pledges to existing or future respondents to its surveys.

19. The decrease in respondent cooperation resulting from a statistical agency's inability to make confidentiality pledges, ensure the integrity of the data, and guarantee equal and timely availability of statistical information to the public, will increase the costs of operating the Federal statistical system. Such continuing increased costs would easily surpass any potential savings resulting from IT consolidation. Also, the quality of the statistical information would decline, which would have very severe implications for decision makers and data users who rely on timely, accurate, and objective information to make informed decisions.

20. In essence, a statistical agency must be distinct from those parts of the department that carry out enforcement and policy-making activities. To be credible, a statistical agency must be impartial and avoid even the appearance that its collection, analysis, and reporting processes might be manipulated for political purposes or that individually identifiable data might be turned over for administrative, regulatory, or enforcement purposes. As noted in *Principles and Practices for a Federal Statistical Agency*:

> ***The authority to ensure that information technology systems fulfill the specialized needs of the statistical agency is another important aspect of independence. A statistical agency must be able to vouch for the integrity, confidentiality, and impartiality of the information collected and maintained under its authority so that it***

[3] "Principles and Practices for a Federal Statistical Agency," fourth edition, National Research Council, National Academy Press, Washington, DC, 2009

> ***retains the trust of its data providers and data users. Such trust is fostered when a statistical agency has control over its information technology resources, and there is no opportunity or perception that policy, program, or regulatory agencies could gain access to records of individual respondents. A statistical agency also needs control over its information technology resources to support timely and accurate release of official statistics, which are often produced under stringent deadlines.[4]***

21. To operate efficiently and effectively, governments, businesses, households, and other organizations depend on timely, objective, and credible statistics. Any loss of trust in the integrity of the Federal statistical system and its products will lessen respondent cooperation with Federal statistical surveys, decrease the quality and increase the costs of statistical system products, and foster uncertainty about the validity of measures our Nation uses to monitor and assess its performance and progress.

22. Fundamental to ensuring continued trust by the public in these principles are requirements on how federal statistical agencies manage their data and information. In doing so, it is important that data confidentiality, integrity, and availability are upheld in a manner consistent with the purposes for which the data and information was collected, in accordance with the *Principles and Practices for a Federal Statistical Agency*, and pursuant to relevant laws governing survey data, such as the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA).

23. As noted in *Principles and Practices for a Federal Statistical Agency*, a large and widely acknowledged position of independence is necessary for a statistical agency to have credibility and to carry out its function to provide an unhindered flow of useful, high-quality information for the public and policy makers. Without the credibility that comes from a strong degree of independence, users may lose trust in the accuracy and objectivity of the agency's data, and data providers may become less willing to cooperate with agency requests.[5]

24. There are similar concerns with respect to commercial data. Terms and conditions for use of commercially available data could also be violated by relocation of statistical agency data storage, potentially subjecting the agency to breach of contract actions. One such vendor—with whom an agency had invested millions of dollars—includes this clause in its terms and conditions for licensing its data:

> "Prohibition Against Non-Specified Use. Any use or disclosure other than as specified in this license is prohibited without [vendor]'s prior written consent. No Information Reference or Software may be … placed on any data retrieval system that may be accessed outside Client's immediate organization…."

---

[4] "*Principles and Practices for a Federal Statistical Agency,*" fourth edition, p. 24.

[5] I would like to acknowledge the contribution of Paul R. Gibson to the Principles and Practices discussion.

VI. Other Vertical vs. Horizontal IT Consolidation Issues

26. *Bigger Stovepipes*: Vertical IT consolidation within parent departments and legal constraints on access limit statistical agencies' ability to horizontally share statistical data among statistical agencies in other departments.

27. *Smaller Budgets*: Resources spent on consolidating IT infrastructure within parent departments are not available to horizontally integrate IT infrastructure across statistical agencies.

28. *Improve Provenance of Data on Data.gov:* By enhancing statistical agencies' linkages to Data.gov, horizontal IT integration through SCOPE would improve data quality and provide more statistical data more quickly, better metadata, and a common platform for better access to online data visualization and analytical tools; ultimately providing end users a single view of Federal statistical data rather than going agency by agency—Data.gov realized.

29. *Efficiencies of Scale Gained Through Sharing:*

- Reduce the number of agencies re-creating similar software tools, analytical models, and collecting similar data
- Improve security through collaboration among agency security experts, implementing best practices, particularly in application development
- Decrease procurement costs through aggregation, open source software

VII. Conclusion

30. Through horizontal IT integration, SCOPE represents the recognition on the part of the leadership of Federal statistical agencies that there are efficiencies to be gained for both the agencies and users by more cross agency collaboration, harmonization of definitions and terminology, identification of best practices, and sharing of the development of common tools that support best practices. ***SCOPE provides an integrated means of identifying mutual challenges and opportunities, then jointly developing and disseminating practical and effective solutions.***