

## SEMINAIRE

## СЕМИНАР

## SEMINAR

STATISTICAL COMMISSION AND  
ECONOMIC COMMISSION FOR EUROPECONFERENCE OF EUROPEAN  
STATISTICIANSDistr.  
GENERALCES/SEM.43/24  
8 February 2000

Original: ENGLISH

Seminar on integrated statistical information  
systems and related matters (ISIS 2000)  
(Riga, Latvia, 29-31 May 2000)

Topic IV: Improving data dissemination strategies

**USER DRIVEN INTEGRATED STATISTICAL SOLUTIONS  
DIGITAL GOVERNMENT BY THE PEOPLE FOR THE PEOPLE**

**Invited paper**Submitted by the U.S. Census Bureau<sup>1</sup>**I. INTRODUCTION**

1. The mission of the U.S. Census Bureau is to be the preeminent collector and provider of timely, relevant, and high quality data about the people and economy of the United States. As we continue to expand the collection and dissemination of data electronically, the Census Bureau has recently undertaken a project, Integrated Statistical Solutions, to explore how we will respond to increasing user requests for integrated data via federal statistical portals such as FedStats. The Integrated Statistical Solutions (ISS) Program articulates a vision and historic opportunity for the federal statistical community and the citizens and taxpayers of our Nation as we enter the 21<sup>st</sup> Century. The ISS Program will implement a modernized, customer-driven, cross-program, and cross-agency integrated data access and dissemination service capability. ISS will broaden information delivery, reduce data user burden, increase efficiencies, and reduce redundancies by providing standards, processes and tools in the administration of information integration metadata repositories; product conception, design, and development; and new disclosure techniques.

**II. BACKGROUND**

2. As we enter this new century amidst a wave of technological innovation, our Nation's statistical community is being confronted with an urgent and critical need for more accurate, timely, relevant, as well as accessible and interpretable data. More powerful computers and rapidly expanding Internet connectivity have revolutionized data access and dissemination capabilities. According to a report by BancAmerica Corporation, there are over 80 million U.S. web users (home and work) today, with projected growth of 20-25% in the coming years. Federal agencies, in particular, face a loss of credibility and, potentially, support if they fail to deliver the public benefits of this new technology to meet rising customer expectations and fulfill the integrated information needs of our Nation and our communities.

---

<sup>1</sup> Prepared by Mark Wallace.

3. These trends are becoming apparent in other countries as well. Sweden, Netherlands, Canada, and Australia are representative of countries where individual statistical agencies are investigating the creation of data and metadata repositories for more strategic use of their statistical data assets including integrated data products. Eurostat also has been long involved in integration and harmonization policies and strategies for national and regional comparative analysis to support social and economic development in the European Union (Tuinen 1994). In the academic community, massive digital libraries are being developed to allow easy access, via the Internet, to multimedia information from a diversity of sources. In the US, FedStats links over 70 federal agencies for easy public access to government produced statistics but has not yet fully addressed the issue of data integration. In fact, the FedStats program and the ISS program, including representatives from several federal agencies, are united in their work with several National Science Foundation Digital Government grantees to fully address the issues required to implement data integration.

4. The ISS Program will support the President's Executive Order 12906 (April 1994) by linking statistical data to geography and time. Executive Order 12906 initiated the creation of a National Spatial Data Infrastructure, a coordinated effort among federal, State, local and tribal governments to support public and private sector applications of geospatial data in such areas as transportation, community development, agriculture, emergency response, environmental management, and information technology. Part of this initiative was to develop a national digital geospatial data framework to support the 2000 Decennial Census. The Census Bureau is in a unique position to add significant value to this initiative with integration of its massive stores of statistical data while at the same time addressing very important confidentiality and disclosure avoidance issues.

5. Recent surveys of Census Bureau customers show that two out of three use multiple data sets. These data users would like to dynamically integrate, manipulate, and examine data from multiple sources and time periods to create enhanced content and awareness. Customers accustomed to surfing the Internet are now expecting the capability to make better decisions via ubiquitous access to rich stores of information. And as we provide increased on-line access to 1997 Economic Census, Census 2000, and American Community Survey (ACS) data in the near future, we will certainly raise expectations even more. If we continue to saddle data users with the burden of compiling, sorting, parsing, reformatting, and otherwise putting data from disparate sources into digestible forms before they can use it, we do it at the risk of our own peril.

6. As more and more data from multiple sources and time periods become available, users are increasingly being left with the additional burden of integrating data sets without the tools or knowledge to know if the data sets can be integrated or the results are meaningful. The ISS Program recognizes that the federal statistical community must work together for the public good to enhance decision-making resources used by our national and local leaders, corporate leaders, small businesses, and citizens. ISS will help minimize data user burden, data uncertainty and maximize data quality and usefulness. ISS will help spur a new data integration industry that will support private industry, the government, and academia. All statistical and geographic data will be fully documented according to agreed-upon standards, tools and resources will be provided to ensure proper usage of Census and other governmental data, and data confidentiality and privacy concerns will be met.

## **II. THE INTEGRATED STATISTICAL SOLUTIONS PROGRAM**

7. The ISS Program is an effort to retool our data creation and delivery systems to deliver "integrated statistical solutions" to customers inside and outside the federal government. It will focus on delivering what customers want, not advancing particular program areas within the government. An "integrated statistical solution" is an answer to a customer's question, delivered without the customer's first having to learn how our programs or our data files are organized.

8. For years, users have relied on federal, state, and local statistics to derive answers to their problems, though this has entailed significant work on their parts to

gather and format the data prior to using it. But, imagine if you will, a customer being able to access a federal web site and to quickly and easily get relevant and useful information on whether it makes sense to re-locate his or her business and family to Anne Arundel County, Maryland. The customer could request data on personally important variables such as retail sales per capita; restaurant sales as a percentage of disposable household income; various age, race, and gender characteristics; cost of living and crime statistics; and recent student-to-teacher ratios and test scores in public schools. Based on this data, compared to other counties in the Washington/Baltimore (or other) metropolitan area, or perhaps counties nationwide if that is what the customer desires, he or she will have valuable information for use in making a good location decision. Or perhaps the user might want to access other available data sets to ascertain other information based solutions such as: How is my community doing? How is my family doing? How is my business doing? What sales targets should I set for this county? Where would be the best places to retire? To realize this level of data service will entail collaboration among several government agencies and their associated databases in the development of an integrated data environment.

9. Our existing data dissemination tools, like American FactFinder and FERRETT at Census and MapStats at FedStats, though very useful in providing flexible access to individual data sets, are not able to provide such a problem and solution oriented view of the data today. The ISS Program, by working in conjunction with FedStats agencies, as well as state and local agencies and the research community, is in the process of extending current capabilities to provide not only individualized streams of products and static data pages but interactive tools, and the establishment of an interagency collaborative development environment, for providing dynamically integrated data and metadata that can help provide information solutions.

#### **IV. THE NEED FOR ISS**

10. The Internet has clearly become the primary distribution channel for data and information. Results of user surveys, focus groups, and ongoing daily contact show the public is demanding broader and quicker access as well as easy point-and-click retrieval and ordering options. Meeting these demands will lead to vastly larger markets and a growing demand for integrated statistical and geographic information solutions. Similar to the effect of Topologically Integrated Geographic Encoding and Referencing (TIGER) on the development of a multi-billion dollar Geographic Information System (GIS) industry, the ISS Program will provide the infrastructure investment that will help spur private sector growth and university research in what will be an emerging data integration industry.

11. The Census Bureau and interagency collaborative efforts such as FedStats have traditionally focused on distributing large tabular data sets (e.g., Summary Tape File data) that contain much more data than many users want. As the focus shifts to information solutions, the user will have access to large numbers of small focused transactions, giving them what they need within privacy and confidentiality constraints. Rather than opening up a fire hydrant of data, it will allow users to take "sips" of only what they need from multiple data sources and time periods.

#### **V. IMPLEMENTATION ISSUES**

12. Data users and providers in the statistical and geographic communities have long been aware of the difficulties in comparing data and/or geography across departments and agencies and/or over time. However, customers have received little support from statistical agencies in their efforts to combine and integrate data from various sources and time periods. As a result, the quality of individual solutions has been highly variable and plagued with data uncertainty. A primary goal of the ISS is to ensure full integration of geographic, demographic and economic data sets. Achieving this capability within the Census Bureau, and with other public agencies via collaborative efforts such as the FedStats Product Concepts Working Group, will require, among other things, the resolution of issues related to definitions of concepts, geography, reference periods, and disclosure avoidance.

## **A. Definitions of Concepts**

13. Data integration is dependent on the use of consistent concepts and definitions across program areas. When programs are administered by different agencies, consistency may be difficult to achieve since program needs may call for different treatments. Many examples exist where data definitions and concepts differ between censuses, surveys, and administrative records and how they change over time. This situation introduces substantial complexities to the storage, processing, display, and documentation of statistical data in an integrated environment.

14. Inconsistency also may result from the lack of coordination, documentation, and communication. Data produced by the Census Bureau and other agencies has historically been the responsibility of specific divisions or branches with little communication with other surveys or programs. This situation has led to stovepipe solutions and corporate inefficiencies. Documentation of these cross-program and cross-agency "discrepancies" will lead to greater standardization in concepts and definitions, a necessary prerequisite for a more robust data integration capability.

15. Data definitions and concepts change over time, particularly for social, demographic, and economic data sets. These situations often present comparability problems. For example, data users are familiar with the ongoing issues and effect of changing concepts and definitions of race and ethnicity over time and varying methodologies for reporting these classifications in censuses, surveys, and administrative records. Another example is the recent replacement of the Standard Industrial Classification system with the North American Industry Classification System and its effect on time series analysis.

16. These definitional and methodological issues demonstrate some of the difficulties and opportunities to develop more consistent policies and standards. They also point to the need for accompanying documentation, including user information, on the "integratability" of various data sets.

## **B. Geography**

17. Geography plays a critical role in any future ISS data integration capability. Data integrity and the quality of information solutions is dependent on a full understanding of the relationship between the geographic "containers" of statistical data, the statistical data itself, and their changes over time. Geographic areas used for reporting often differ across programs and between agencies. Comparing data from a specific survey or census between two time periods as well as between different censuses and surveys with different reference dates are frequent obstacles for robust data integration. The relative timing of updates to the geographic data base (e.g., boundary changes) vis a vis the reference period of demographic and economic data collection also presents issues that must be addressed.

18. Different surveys use different geographic frames; that is, the universe of available geographic entities for which data are reported may differ from one data set to another. In some cases, the universe of legally existing geographic entities may be modified for statistical reporting purposes. In other cases, data may not be reported for some entities at a specified level of geography. Another situation affecting data integration is the "longitudinal" approach used by some demographic surveys in reporting current data with older versions of geography (e.g., metropolitan areas).

19. Different reference dates used for collecting and reporting data create uncertainty on the extent of differences between data elements from one data set to another. Demographic and economic data may be collected as of a specified date (i.e., April 2000) or, as in the case of administrative records, over a period of time. Even if there are no changes in geography, different reference dates may pose obstacles to data integration, especially for areas experiencing rapid changes in demographics.

20. Demographic, economic, and geographic changes occur across time and space. Capturing snapshots over time will always present some obstacles because of the different reference periods used to collect and present data, availability of data for

specific types of geographic entities across agencies and time, and changes in spatial data. These obstacles to data integration are more pronounced at lower levels of geography. Solutions to these issues, and the ability to re-use and re-purpose data, will revolve around the further development of standards, metadata, an enhanced geographic information processing environment, and a respondent/location level data capability.

### **C. Disclosure Avoidance**

21. The protection against inadvertent disclosure of individual persons, households, establishments or primary sampling units, especially in public use databases, is a foremost concern of government statistical agencies. While georeferenced identifiers can greatly increase the value of data analysis, they increase the government's burden to ensure the protection of the individual's right to confidentiality. As such, the ability to integrate data sets is dependent on the resolution of privacy and confidentiality issues, the development of sophisticated disclosure avoidance methodologies, and newly devised data presentation methods.

22. Currently, published data from Bureau pre-defined tables that have already passed disclosure avoidance review are suitable for integration. Disclosure problems arise when attempts are made either to make the geography of the different data sources the same or to publish all of the data from the different data sources for the same geography.

23. While query systems add additional impetus to the long standing desire to have comparable geography and definitions between surveys, the problem remains that geography was often chosen to provide maximum information from a survey to its primary users given the usual constraints of sample size and disclosure. Imposing geography driven by another survey is likely to violate at least one of these considerations. A query system that re-tabulates the underlying survey data with modified geography for comparability purposes would need to guard against disclosure.

24. In conjunction with research efforts under the auspices of the Digital Government Initiative, the ISS will continue to investigate these issues and develop solutions, to the extent possible. County-level solutions may be easier to implement as it is a unit already common to many surveys. Cell suppression technology could offer some solutions to these obstacles. Currently, tests of the current data access and dissemination system query system are being done for lower levels of geography from the Census 2000 dress rehearsal data with checks for complementary disclosure.

## **VI. IMPLEMENTATION STRATEGIES**

25. The Integrated Statistical Solutions (ISS) Program will use and promote a corporate approach to project management and a collaborative approach with other government agencies and the research community as part of the implementation strategy. Working with other agencies via the FedStats Product Concepts Working Group, it will evolve from existing programs within the government such as DADS, FERRETT, Centers for Disease Control "(CDC) Wonder", and MapStats in prototyping to prove in new concepts and technologies. ISS will systematically seek and integrate internal and external user input and will develop partnerships with other agencies and the research community in designing, developing, and reviewing prototypes of next generation data access tools that can be implemented via new and developing technology solutions. Moreover, the establishment of metadata repositories, logical data warehouses, standards and business rules and practices, and customer relationship management systems, will comprise the centerpiece of ISS.

### **A. Metadata**

26. Implementation of ISS requires development of a corporate metadata repository (CMR) along with the development of tools to create and add metadata and extract it from the repository. Metadata, or data about data, is the descriptive information which describes data content, organization, derivation and limitations (Gillman, 1994). Without electronic metadata repositories such as distributed dissemination

metadata repositories (DDMRs) associated with CMRs in the various statistical agencies, the ability of ISS to develop integrated information is seriously constrained.

27. Documenting data sets is a first step in reengineering data collection and dissemination, and enabling data sharing between agencies. Statistical and geographic databases have been built thus far to support the mandates of single institutions or parts of an institution. All who collect and manage data for activities related to their own responsibilities will need to understand and appreciate the value of those data to others and to collect and structure their data accordingly. To take full advantage of the opportunities offered by new technologies, business, government, and academia will need to develop, support, and fund metadata on a systematic and ongoing basis as well as promote access for all.

28. The Internet has clearly changed expectations and heightened knowledge about the ease of access to information as well as broadened the universe of users. Customers, both internal and external, associated with the various government agencies expect technology to provide easy and rapid access to information and data. The expectation is for better documented, usable and interpretable data. Developing CMRs and DDMRs will enable us to address the increasing demand for rapid access to documented and usable data by collecting business information across the full survey/census life cycle. Metadata repository technology, as implemented in a loosely coupled, virtual way, is scalable. Local repositories will store metadata and provide the ability to share and reuse metadata on an enterprise-wide level. This might eventually be expandable to a virtual government-wide, or even global, level. Metadata will be the integrating mechanism for providing government-wide information.

29. Presently the Bureau's business processes touch metadata from the survey/census design through the dissemination life cycle. Yet, metadata is not formally cataloged until the dissemination phase where it is required for Bureau dissemination systems and products. The ISS approach to metadata will be to catalog these data at each phase of the survey life cycle so that metadata does not have to be recreated during the dissemination phase. This approach also provides value added during each of the core business processes.

## **B. Standards and Business Practices**

30. The ISS Program will encourage the Bureau-wide use of common processes for product conception, development and delivery including integration of customer feedback; re-use of existing data sets to reduce respondent burden; easy electronic access to all data and metadata sources. Standards will be developed and existing standards promoted across the enterprise for data integration, product quality, usability, archiving practices and project management. Scalable templates will be available for business cases, marketing plans, product testing procedures, and change control procedures along with guidance. ISS also will support and promote the use of corporate business practices for the use of non-Census data for integrated products, security practices and disclosure avoidance issues.

31. In the ISS Program vision, improvements are shared corporately as they are realized locally, resulting in agency efficiencies, lower costs, less duplication of effort, and shared corporate-wide quality standards. Building ISS around the aforementioned business practices will support a better integration of the organization and technology. Standardized development and access will promote sustainable and repeatable processes less dependent on the ad hoc knowledge and expertise of a few individuals. It also will facilitate the development of a common approach to customers, products, and resources.

## **C. Data Warehouse**

32. The data warehouse provides access to numerous data sets that have been consolidated and optimized for decision support purposes. A hybrid model (incorporating elements of both centralized and distributed technology) appears to be the best model for satisfying Census requirements, while meeting organizational

constraints. In addition, the hybrid model has the ability to accommodate any distribution alternative, from fully centralized to fully distributed.

33. The hybrid architecture consists of a logical warehouse engine, based on data integration tools, with an application programming interface (API) for application connection and a user interface for user query and extraction purposes. It is flexible enough to support all data in the primary warehouse or, alternatively, to place all data in tertiary databases. Most importantly, the architecture supports an any-to-any migration, providing a highly flexible modular environment. As a result, the modularity of the architecture will enable the Bureau to re-evaluate implementation decisions, and to alter the implementation as new policy decisions or technologies emerge, demand for specific data sets changes, or infrastructure improvements are made.

34. The architecture supports the storage of data in primary, secondary, and tertiary databases:

- primary warehouse (analogous to today's American FactFinder system);
- secondary databases, other database instances either on the same server, or on another server on the same high speed local area network; and
- tertiary databases, other databases running on servers on other locations, with three levels distinguished by the speed of the connecting infrastructure

35. Data and the associated metadata from other systems may be connected through the logical warehouse technology if it is already transformed to a warehouse usable format. If not, it must be transformed, and may be placed in a primary, secondary or tertiary database for use by the logical warehouse. (Metadata may reside locally to improve performance, in the case of in-house or remote data sets.)

36. The data integration technology will allow the joining of different data sets across platforms. For example, it will allow users to query many different data sets across different platforms natively. This will foster an environment in which data and metadata set owners will be responsible for maintaining their own data sets while enabling a single point of access to all data sets. This eliminates much of the complexity of connecting to new data sets on unfamiliar platforms, and eliminates the need for all the data and metadata included in the system to be maintained by one organization -- which would likely create a system bottleneck. In addition, it provides an opportunity to revisit the publication processes and tools in use today. There may be an opportunity for significant gains in efficiency and effectiveness of the process, simplifying the ability to target multiple media types with the same product.

#### **D. Supporting Spatial and Statistical Infrastructure**

37. One of the primary assets for integrating Bureau-wide and government data sets is the Census Bureau's TIGER database. The creation of this nationwide digital geographic database for the 1990 census, provided a major impetus for GIS developments in the last decade and the ability to easily overlay statistical and other data for common geographic entities. The ISS will extend these developments and create a robust spatial/statistical data integration industry that will support the National Information Infrastructure and National Spatial Data Infrastructure that will feed national and local decision support systems.

38. The MAF/TIGER database supports the ISS vision of moving from an organization of division-specific providers of traditionally separate data products fixed in time and space to a Bureau-wide provider of continuous information solutions using the most current and reliable data and visualization technology available. TIGER, a nation-wide and public digital geographic database that links all the Census Bureau's extensive social, demographic, and economic data sets and the MAF, a nationwide inventory of all residential, and eventually, commercial units are key components of the Census Bureau's data infrastructure that will enable linking, integrating, and visualizing Bureau-wide and government-wide data sets.

## **E. Interagency Collaborative Projects**

39. Various teams composed of staff from all directorates of the Census Bureau in collaboration with the State Data Centers, FedStats agencies, and the research community is researching the potential and validating data integration processes for accessing and integrating both micro and macro level datasets (and their metadata) that are Census Bureau-based and/or remote. These teams also are developing data integration tools to create profiles (predefined and user-specified), and to give customers new data integration functionality, providing information based solutions not currently available in our existing data access tools (AFF, FERRETT, MapStats, etc.). Through work on these teams, we will address many of the aforementioned issues connected with developing Integrated Statistical Solutions.

40. Major activities within the scope of the project mission and which are the responsibility of the various project teams are to:

- Identify datasets and data items for use in pilots.
- Specify the content (e.g., datasets, data items, data combinations) of the integrated data tools and products (integrated statistical profiles - statistical abstracts of federal data at sub-national levels of geography; "Hot Reports" - a merging of text, graphics, and maps - along with the data for integrated statistical profiles, and other new tools and products).
- Specify the functional requirements for user-defined profiles.
- Employ data warehouse, CMR, and DDMR concepts, research processes for accessing and integrating both micro and macro level datasets (and their metadata) that are Census Bureau-based and/or remote.
- Build, test, and document a prototype of the data access and integration process.
- Design a UI that specifies input screens and result screens for both predefined and user-specified profiles.
- Build, test, and document a prototype of the UI.
- Design an application to create profiles (predefined and user-specified), and to give data users new functionality not currently available in our existing data access tools.
- Build, test, and document a prototype of the application.

41. These projects will make use of a collaborative, multi-agency laboratory server environment that is being established to evaluate different tools and approaches, use corporate metadata policies and techniques, and include profiles with some level of graphic display capabilities. As a result of this work, the ISS Program will begin to:

- ▶ Define what kind of data profiles users need (including the relationship between predefined and user-defined profiles, and what customers want the system to do).
- ▶ Address geography, time, data item, and other comparability/integrateability constraints.
- ▶ Determine criteria for inclusion of datasets, including the availability of metadata; geography supported; comparability of geography, time, and/or other definitions; disclosure avoidance constraints; and constraints imposed by data partners.
- ▶ Identify datasets, items, and combinations to be included.
- ▶ Validate use scenarios with users.

## **F. Project Management**

42. The Census Bureau has recently embarked upon an enterprise-wide effort to improve the quality and standardize the practice of project management by instituting a Project Management Certificate program developed by ESI International in association with George Washington University. The basis for the program is the Project Management Body of Knowledge (PMBOK) developed by the internationally recognized Project Management Institute (PMI) to provide a consistent structure for project management professional and development programs including: (1) Certification of



Project Management Professionals (PMPs) and (2) Accreditation of degree-granting educational programs in project management.

43. Implementation of the ISS Program, including pilot projects, will be based on the new project management approach, will utilize the practices and tools, and will promote their use as the Bureau standard. Key personnel already involved in ISS activities include a cadre of certified project managers who have completed the certificate program. ISS will be one of the first programs to reap the benefits of this program and will further promote the use of the corporate approach to project management by building this standard into the processes for conception, development and delivery of new integrated data products and tools. This synergistic relationship is one of many the ISS Program will be able to promote.

## VII. CONCLUSION

44. Converging trends in technology, databases, customer expectations, and societal needs present the Nation's statistical reporting infrastructure a unique opportunity to achieve a quantum leap of excellence in government service. With its critical mass of data, core competencies, an integrated statistical and geographic processing environment, and a massive data sharing infrastructure, the U.S. federal statistical agencies are uniquely positioned to collaborate with local statistical agencies to take advantage of these emerging trends. In the process, it can provide a national leadership role in quality-driven integrated dissemination that will feed many of our Nation's critical decision support systems. In summary, the ISS Program is good government that makes sense.

45. The ISS Program is a next-generation national statistical information infrastructure investment that will modernize the delivery of information services and improve data integration tools and products. The ISS Program will enable federal statistical agencies to meet emerging internal and external customer demands for more accurate, timely, and relevant data and do so in a manner that saves taxpayer dollars, improves data quality, relieves data user burden, and creates an environment for enabling synergistic linkages with ongoing initiatives at agencies like the Census Bureau and other federal, state, and local statistical agencies. Providing integrated statistical solutions requires the ability to combine public resources to maximize use and re-use of our Nation's data resources. The ISS Program is currently investigating several issues and establishing multiple partnerships that will ensure the full benefits the ISS Program for the Census Bureau and the government, in general.

46. The ability to integrate, overlay, and visualize data and geography is increasingly critical for our national, regional, and local decision support systems in fields such as transportation, community development, agriculture, emergency response, public health, environmental management, and information technology. The ability to provide current and accurate data that can be quickly integrated with diverse government-wide data sets is critical to the Nation's economy and is a means to provide a better quality of life for all.

## REFERENCES

- Clinton, William J., (1994), Executive Order 12906, Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure, Washington, DC: The White House, Office of the Press Secretary.
- Duncan, William R., PMI Standards Committee, (1996), A Guide to the Project Management Body of Knowledge, Upper Darby, PA: Project Management Institute.
- Gillman, Daniel W. and Appel, Martin V., (1994), "Metadata Database Development at the Census Bureau", paper presented at the UN/ECE Metadata workshop, Geneva, Switzerland.
- IBM Consulting Group, (1999), Integrated Statistical Solutions (ISS) Technology Roadmap, IBM Consulting Group.

- Sundgren, Bo, (1998), An Information System Architecture for National and International Statistical Organizations, an invited paper prepared for the Conference of European Statistics, Meeting on the Management of Statistical Information Technology, Geneva, Switzerland.
- Tuinen, H.K. van, J. W. Altena and H.C.M. Imben (1994), "Surveys, Registers and Integration in Social Statistics, Statistical Journal of the United States Economic Commission of Europe, pp. 321-356.
- Wallace, Mark E., Landman, Cheryl M., Sperling, Jon, and Buczinski, Carla, (1999) "Integrated Information Solutions - The Future of Census Bureau Data Access and Dissemination.