

SEMINAIRE

СЕМИНАР

SEMINAR

STATISTICAL COMMISSION AND
ECONOMIC COMMISSION FOR EUROPECONFERENCE OF EUROPEAN
STATISTICIANSDistr.
GENERALCES/SEM.43/24 (Summary)
2 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 paperSubmitted by the U.S. Census Bureau¹**SUMMARY**

1. We are well into the information age. The Internet has changed everything. Technology and user expectations have rapidly advanced to the point where government dissemination can and must provide answers to user questions - not just provide statistics. The provision of these answers will be accomplished by successfully integrating data files from many sources to help provide statistical solutions for data users both inside and outside statistical agencies. Increasing the public availability of government information and increasing statistical services to the citizen will be accomplished via the Integrated Statistical Solutions (ISS) initiative. The U.S. Census Bureau, with its data, technology, and user service expertise, is in a most advantageous position to meet the American public's emerging information needs - to move beyond mere data dissemination to integrated statistical dissemination.
2. The U.S. Census Bureau instituted the ISS Project to define the details of how we should proceed from the development of the Data Access and Dissemination System (DADS) project to the full vision of an integrated information access and dissemination system from an external data user and an internal processing perspective.
3. If implemented in collaboration with other government agencies, the ISS Program will produce a modernized, customer-driven, cross-program, and cross-agency integrated data access and dissemination service capability available via statistical portals such as FedStats. ISS will broaden information delivery, reduce data user burden, increase efficiency, and reduce redundancies by providing standards, processes and tools in the administration of information integration architectures; metadata repositories; product conception, design, and development; and new disclosure techniques. ISS will serve as a model and catalyst for change in the federal statistical reporting community. Moreover, it will help build critical capabilities

¹ Prepared by Mark Wallace.

in the Nation's emerging statistical and spatial data infrastructures that will support global, national, regional, local, and individual decision support systems. This paper describes the ISS Program goals to retool product concept, creation and delivery systems to deliver "integrated statistical solutions" to customers inside and outside the federal government.

Implementation Issues and Strategies

4. 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 with government agencies and the research community will require among other things, the resolution of issues related to definitions of concepts, geography, reference periods, and disclosure avoidance.

5. The paper will discuss pilot projects for integrating data sets in terms of an overall implementation strategy for the program. These pilots include utilizing the data warehousing capabilities of DADS along with the more aggressive remote access capabilities of the Federal Electronic Research and Review Extraction Tool (FERRETT), also being developed at the U.S. Census Bureau, to link Census 2000 data (and socio-economic data from elsewhere in the Census Bureau) to other federal, state, and local data sources.

6. ISS will evolve from existing programs within the government such as DADS, FERRETT, Centers for Disease Control "(CDC) Wonder", and FedStats' 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.

7. 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. Developing corporate metadata repositories (CMRs) and distributed dissemination metadata repositories (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. Local repositories will store metadata and provide the ability to share and reuse metadata on an enterprise-wide level and eventually government-wide. Metadata will become the integrating mechanism for providing government-wide information.

8. 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.

9. A **data warehouse** will provide 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 approach. 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.

10. One of the primary assets for integrating Bureau-wide and government data sets is the Census Bureau's TIGER database. The creation of this nation-wide 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.

Interagency Collaborative Projects

11. Various teams composed of staff from all directorates of the Census Bureau in collaboration with the State Data Centers, the FedStats Product Concepts Working Group, and the research community are researching the potential and validating data integration processes for accessing and integrating both micro and macro level data sets (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. Through work on these teams, we will address many of the issues connected with developing Integrated Statistical Solutions.

12. These collaborative efforts will make use of a 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 data sets, 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 data sets, items, and combinations to be included;
- ▶ Validate use scenarios with users.

13. 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.

Conclusions

14. 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.