

Distr.  
GENERAL

CES/AC.71/2005/12 (Summary)  
31 January 2005

Original: ENGLISH

**UNITED NATIONS STATISTICAL COMMISSION and  
ECONOMIC COMMISSION FOR EUROPE (ECE)  
CONFERENCE OF EUROPEAN STATISTICIANS**

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

**ORGANISATION FOR ECONOMIC  
COOPERATION AND DEVELOPMENT (OECD)  
STATISTICS DIRECTORATE**

**Joint ECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS)**  
(Bratislava, Slovakia, 18-20 April 2005)

Topic (ii): Development strategies for statistical information systems

**TOWARDS A UNIFIED PUBLICATIONS SYSTEM AT THE  
U.S. BUREAU OF LABOR STATISTICS**

**Invited Paper**

Submitted by the Bureau of Labor Statistics, USA<sup>1</sup>

**Summary**

**I. INTRODUCTION**

1. Once upon a time the publications process at the U.S. Bureau of Labor Statistics (BLS) was relatively straightforward. Program offices would prepare their materials; these would be submitted to the Office of Publications for review; from there they would be sent out for typesetting and printing; and finally the finished publication would be distributed to interested readers – primarily through the postal service. Desktop publishing capabilities moved more of the physical production process in-house but the basic flow remained largely intact.
2. The advent of the World Wide Web and other electronic data dissemination methods complicated this picture quite a bit. Not only were there now two different groups receiving final text (one for print publications and one for on-line distribution) but each group required different formats: camera-ready copy vs. HTML, respectively. What's more, customers quickly demanded more than preformatted tables and centrally established data presentations. Public Web users wanted to select their own data series and arrange the estimates according to their individual requirements. Thus survey data also needed to be available in an on-line database accompanied by a family of end-user query and formatting tools.
3. These changes took place over a span of years and nobody knew at the outset just where they would lead. Thus the output modules of survey production systems were tweaked to accommodate an incremental progression of enhancements. The result is a conglomeration of patchwork systems, one per program.

---

<sup>1</sup> Prepared by Michael D. Levi (levi.michael@bls.gov).

4. This paper will describe the steps the U.S. Bureau of Labor Statistics is taking to analyze and build a unified publications system that will serve all program areas and output channels.

## **II. BACKGROUND**

5. Currently the following channels are used to disseminate BLS data to end users: print (news releases, detailed reports, The Monthly Labor Review and other dedicated journals), the Internet (preformatted Web pages, a collection of interactive query tools that access a central data repository, flat files and Excel spreadsheets via FTP, and an e-mail subscription service), automated FAX-on-Demand, and a telephone-based interactive voice recognition system. On the drawing board is wireless transmission to personal digital assistants, cell phone, and other handheld devices.

6. The most significant factor behind the current collection of systems is a history of independent development conducted by largely autonomous program areas.

7. In 2003 the BLS Internet division released its Table Generation System (TGS), which pulls estimates from the central database and generates statistical tables formatted to comply with U.S. government regulations for accessibility to people with disabilities (Section 502 compliance.)

## **III. PROBLEMS WITH THE CURRENT SYSTEM**

8. There are two levels of duplication in the current environment: not only does each program have its own statistical output subsystem, but within each program there are parallel output modules for on-line and hard-copy dissemination.

9. Accompanying the within-program redundancy described above is a risk of inconsistent data reaching the public. Though rare, it is not unheard-of for a last-minute change to be implemented in one output stream but not in the other. The time lag between the publication of conflicting data and its identification and correction is an embarrassment at best and a disservice to our customers at worst.

10. Accompanying the cross-program redundancy described above is an unnecessarily slow response to full agency implementation of new technological opportunities.

11. Finally, the current system lacks unifying within-program and cross-program metadata that might be exploited to improve end-user access and comprehension.

## **IV. VISION FOR NEW SYSTEM**

12. The key elements of a replacement publications system that will serve the needs of the agency as well as the customer base are a central database containing all published data; an XML schema that describes statistical tables and publications formats (including explanatory or analytic text); and a set of transformation routines that will produce the required set of output products.

13. Once the basic functionality has been established, numerous expansions that go far beyond current capabilities become possible. Some examples include enhanced Web search capabilities; automatic generation of cross-references and suggestions for related materials; full graphing, mapping, and other data visualizations; and greatly improved cross-program end-user tools.

## **V. BARRIERS TO IMPLEMENTATION**

14. Not all programs are convinced of the need for a new system. Some of their objections include:
- "If it ain't broke don't fix it." The current approach successfully distributes BLS information to the public month after month, thus there is reduced urgency for change;
  - A substantial development effort will place additional (short-term) demands on program staff and will interfere with other high-priority projects;
  - Programs anticipate a reduction in control over the products that define their success and are concerned that their idiosyncratic requirements may not be fully implemented;
  - There is no guarantee that publication-quality output can be generated automatically without extensive hand tweaking.

## **VI. PROGRESS TO DATE**

15. Several of the essential components of a new system have been developed and successfully implemented. In particular, a central database containing the vast majority of published data is in place, a core XML schema for statistical tables has been developed and is in use, and the first set of transformations (XML plus data to HTML) has been running for eighteen months.

16. An unexpected opportunity recently opened: two major programs are rewriting their IT systems from scratch and have recognized the efficiencies an externally developed and operated system could provide. These two programs have jointly chartered an exploratory team to analyze and write specifications for a full replacement publications subsystem; the Office of Publications and the Internet development and operations group are full partners in this team.

17. The author of this paper expects these two programs to provide the impetus and organizational weight necessary to bring the publications project to fruition, with full release expected in 2008.

-----