Dissemination Of Statistical Products: The IRS's Journey

by Martha Eller Gangi, Internal Revenue Service

With the Revenue Act of 1916, the U.S. Congress required that statistics of federal income taxation be prepared and published on an annual basis. In 1917, the predecessor of today's Statistics of Income (SOI), a division of the U.S. Internal Revenue Service (IRS), was created to meet that requirement. The division, the primary source of data on the U.S. tax system and one component of the decentralized U.S. statistical system, has published print reports on federal income since 1918, when the first report on personal and corporate income taxes was released. Many decades later, in 1981, SOI began publishing the *Statistics of Income Bulletin*, a quarterly print report that provides the most current financial statistics from various types of tax and information returns, as well as information from special analytical studies of interest to students of the U.S. tax system, tax policymakers and tax administrators. In addition, the division also produces separate annual reports on individual and corporation income tax returns that contain more comprehensive data than those published in the *Statistics of Income Bulletin*. In 1996, after four years of disseminating data via an electronic bulletin board, SOI introduced its "Tax Stats" pages on the IRS web site, irs.gov.

Today, SOI continues to publish several print products, including its quarterly *Statistics of Income Bulletin*, which presents between four and eight articles per issue, most of which are written by SOI authors; the *IRS Data Book*, the equivalent of an IRS annual report; the *IRS Research Bulletin*, a compilation of tax research papers by U.S. and foreign authors presented at the annual IRS Research Conference; the SOI Paper Series, articles presented by SOI staff and other researchers at professional conferences in the U.S. and abroad; and annual reports on individual and corporation income tax returns. The Tax Stats web site offers all of SOI's print products in electronic format, as well as more than 7,600 statistical files and tables. The site averages approximately 300,000 to 400,000 downloads per month, with nearly 12,000 regular subscribers to its e-mail dispatch system.

The journey from a predominantly print-centric publishing environment, in which web dissemination is an afterthought, to an environment in which at least equal attention is given to print and web dissemination, with the web taking precedence in many cases, is a challenging one. This paper will document SOI's journey from print-centric dissemination of federal tax statistics to a hybrid of print and web dissemination. Specifically, the paper will discuss recent efforts to re-evaluate and improve SOI's print and web products; create and publish internal guidelines for producing user-friendly statistical tables for the web and print; and make SOI's web products more accessible to a broader audience, including those with vision and mobility impairments.

▶ The Age Of Electronic Dissemination

The Birth of SOI's Tax Stats Web Site

For most of its history, SOI relied on print products to disseminate statistics of federal income taxation. However, in 1992, almost 80 years after its inception, the division began to disseminate statistical products via an electronic bulletin board. SOI's age of electronic dissemination had dawned. After only four years of disseminating data via bulletin board, the division gained a presence on IRS's web site, irs.gov. The first Tax Stats web pages offered approximately 700 statistical products, including electronic versions of SOI's publications, such as the *Statistics of Income Bulletin*, and data tables prepared primarily for print dissemination.

This paper was originally presented at the United Nations Statistical Commission and Economic Commission for Europe Work Session on Statistical Dissemination and Communication on May 13-15, 2008.

In the years immediately following the birth of SOI's Tax Stats web site, the number of SOI products and services offered via web increased. By 2003, more than 1,700 files were made available to SOI customers. While web content and downloads from the site increased annually, a satisfaction survey distributed to customers who contacted SOI's information office, members of the U.S. National Tax Association and tax policy researchers who serve as advisors to SOI revealed that SOI's products and services, although useful and of high quality, were difficult to locate on Tax Stats. This survey, administered in 2003, was the first of several steps taken to rethink and redesign SOI's Tax Stats web site. A Tax Stats Web Advisory Group comprised of senior economists, statisticians and computer specialists from a wide range of subject-matter areas across the division was organized to learn more about customer experiences on Tax Stats, address design and content problems with the site and guide future development of Tax Stats pages, within limits imposed by the parent site, irs.gov, which exists predominantly for tax administration.[1]

Evaluation of SOI Products

While the Tax Stats Web Advisory Group was formed to address SOI's web needs, a second group was formed to examine and evaluate the numerous data tables and publications produced within the division and made available via print and web. The Workgroup for the Evaluation of SOI Tables and Publications (WESTP), comprised of SOI senior staff, primarily frequent contributors to SOI publications, and three tax policy researchers who serve as advisors to SOI, was charged with increasing the usability and accessibility of SOI products.

Faced with many opportunities to improve usability and accessibility of tables and publications, WESTP members chose to first address SOI's most visible publication, the quarterly *SOI Bulletin*, and the process by which it was produced and disseminated. The group examined the publications of several other statistical agencies in the U.S. and discussed potential application of their best features, including short articles on topics of current interest and trend analysis, to the *Bulletin*. Based on their own experiences, the group recognized that tables included in the historic data section of the *Bulletin* also needed revision, particularly when uploaded to Tax Stats. Ultimately, the group turned to two methods of information-gathering to focus their efforts. The group administered a satisfaction survey to better understand how SOI customers used the *Bulletin*, how satisfied customers were with the content, how useful the various features of the publication were to them and how it should be improved. In addition, the group held focus groups with *Bulletin* authors and in-house technical reviewers to solicit their feedback on the production process.

Improvements to SOI Data Tables

While almost 90 percent of those who responded to the *Bulletin* survey were satisfied or totally satisfied with the *Bulletin*, areas for improvement were identified.[2] Nearly half of all respondents indicated that they would like to see more articles on topics of current interest, and more than a third said that they would like the *Bulletin* to include shorter articles. The majority of respondents also indicated that they prefer tables and articles to be made available in both print and web format, rather than one or the other. A third of survey participants chose to respond to the open-ended question, "If you could change one thing about the *Bulletin,* what would it be?" Those participants requested more topical, interesting articles; provision of links to related, technical documentation on SOI's Tax Stats web site;

[1] For a thorough discussion of Tax Stats' redesign and the challenge of meeting customer needs for statistical products on a site predominantly designed for tax administration, see Johnson, Barry W. (2006), "Standing Out in a Crowd: Improving Customer Utility on a Centrally Administered, Shared Web Site," *United Nations Economic Commission for Europe, Work Session on Statistical Dissemination and Communication*, available at www.unece.org/stats/documents/2006.09.dissemination.htm.

[2] For more detailed results from SOI's survey of *Bulletin* customers, see Schwartz, Ruth and Beth Kilss (2006), "Customer Satisfaction Initiatives at IRS's Statistics of Income: Using Surveys to Improve Customer Service," *2006 Proceedings of the American Statistical Association*, Statistical Computing Section [CD-ROM], Alexandria, VA: American Statistical Association.

and creation of electronically usable *Bulletin* tables for upload to Tax Stats.

The *Bulletin* survey confirmed WESTP members' own instincts about the need for more usable and accessible data tables on the web, which were created in Microsoft Excel primarily for print publication and then uploaded to Tax Stats with little or no modification. For example, data tables produced for print included notations for cell suppression and other statistical purposes, and such formatting prohibited simple mathematical analysis of table values when uploaded to the web. By applying notations without using Excel's cell-formatting function, cell contents were transformed from arithmetic values to non-arithmetic values that limited use by SOI customers. Table titles and footnotes prepared for print publication were also problematic when tables were simply uploaded to Tax Stats, as both disappeared from view when customers scrolled through tables that, in many cases, contained numerous columns and rows.

To address these issues, a few members of WESTP developed pilot Excel data tables for distribution to, and testing by, a small number of SOI advisors. The pilot tables included only arithmetic values in body cells, and table titles and footnotes were each included in one row with stub titles frozen for continuous viewing of titles, footnotes and stub titles. Page and column breaks were also removed from pilot tables to accommodate standard pages. Advisors responded positively to the change and were pleased with the ability to perform mathematical analysis without reformatting. Pilot Excel tables were the first step in creating guidelines to produce tables for use in print and then to upload to the web. WESTP solicited feedback from SOI economists and others who utilized the early guidelines, and revisions were made with the recognition that the guidelines would continue to be a "living" document—ready for update as necessary and as requirements for accessibility by those with vision and mobility impairments were better understood.[3]

Creating a Streamlined Production Process

The second information-gathering method utilized by WESTP, focus groups with *Bulletin* authors and technical reviewers, revealed that, on the whole, authors and reviewers were very satisfied with the final product—the print version of the *Bulletin*. However, results suggested that the production process itself, as well as the process for getting articles and tables to the web, needed to be improved. In addition, getting usable Excel data tables to the web as early as possible, even before making them available via print, where possible, was identified as a priority.

Under the original production process, SOI's publications team—two visual design specialists, a writer-editor and a publishing services coordinator—served as the intermediary between *Bulletin* authors and reviewers and essentially steered articles through every part of the production process, from creation of customized production schedules for each author to oversight of the necessary layers of article and table review, including technical reviews, quality reviews and reviews by outside experts at the U.S. Treasury Department's Office of Tax Analysis. The back and forth between authors, reviewers and the publications team was time-consuming and increased the likelihood that errors could be introduced.

The publications team and WESTP members analyzed focus group results and examined each step of the production process—from article conception to publication. In the summer of 2006, a revised process was introduced. Under the new process, the publications team and authors, with input from their subject-matter areas, would agree upon a due date for layout, and authors and their areas would be responsible for completing the necessary layers of article and table review by the due date. As soon as data were available and as appropriate, data tables would be produced following the new guidelines and uploaded to Tax Stats, and those web-ready tables would then be submitted to the pub-

[3] For further information on the requirements for accessibility, see Section 508 of the Rehabilitation Act (29 U.S.C. 794d), as amended by the Workforce Investment Act of 1998 (P.L. 105-220), August 7, 1998.

lications team for use in the layout process. The publications team would produce publication-quality tables and figures, lay out each article and conduct editorial and final reviews, working iteratively with authors, as necessary. Soon after its introduction, the new process yielded benefits, as production times decreased almost two weeks, and getting data to the web as soon as possible, rather than waiting for a print product, has allowed earlier access by the public.

▶ Near Parity: Statistical Dissemination Via Print And Web

Recent Initiatives in Statistical Dissemination

By 2007, SOI leadership, those responsible for SOI publications and the Tax Stats web site, and subject-matter experts had begun to understand that dissemination of SOI data was a multiple-stage process, in most cases. Data are uploaded to Tax Stats in Excel table format as soon as they are available, then made available via print and finally made available through Adobe Acrobat PDF versions of print products, also uploaded to the web. Neither means of dissemination—web or print—was seen as dominant. Files available on Tax Stats grew to approximately 6,600 and print publications were still in demand, although the number of copies necessary to meet demand had dropped slightly, perhaps due to availability via web. SOI's publications team added two visual design specialists, which allowed the team to turn its attention to the next challenge—satisfying requirements for accessibility outlined in legislation which mandated that federal agencies provide comparable access to, and use of, information and data to all members of the public, those with and without vision and mobility impairments.[4]

In mid-2007, SOI hired a consultant/designer to recommend a proposed layout process for electronic products that would accommodate compliance with accessibility mandates, as well as technical requirements for the printing and graphic arts industry. The consultant made recommendations to improve the appearance and readability of the *Bulletin*, to enhance conversion to compliant PDF files suitable for distribution via the web, and to make Excel tables on the web accessible to those with vision and mobility impairments. The publications team steadily began to incorporate recommendations for the *Bulletin's* layout process, including use of OpenType fonts for cross-platform compatibility; use of paragraph and character styles; use of Microsoft Word's automatically generated footnotes; and use of text threading and linked frames in desktop publishing software, InDesign, to maintain reading flow in PDFs uploaded to the web. In addition, these recommendations were also applied to the layout process for other SOI publications, including the *IRS Data Book* and the *IRS Research Bulletin*. The publications team worked with several former members of WESTP, which disbanded in late 2006, to include additional accessibility requirements in SOI's table guidelines. In late 2007, the new guidelines were applied to historic data tables and other tables available on Tax Stats.

Change and innovation were contagious, and other initiatives in dissemination of statistical products followed. Also in 2007, SOI introduced the SOI Paper Series, which replaced *Special Studies in Federal Tax Statistics,* a compilation of articles presented by SOI staff and other researchers at professional conferences in the U.S. and abroad. Since few customers were interested in all of the articles in the publication, and most articles were also available on the web, demand for the entire print publication had decreased significantly. Now, bound print versions of individual SOI Papers are produced in-house and made available, at customer request.

Also in 2007, SOI implemented use of credit and debit cards for the purchase of SOI reimbursable products, including recent zip code and migration data, and special tabulations, among others. By early 2008, almost two-thirds of all product sales through SOI's information office were completed via secured web

[4] For a summary of accessibility issues that face federal statistical agencies, see Brown, Laurie and Marianne Zawitz (2004), "Issues for Statistical Agencies: Implementing Section 508 on Agency Web Sites," FedStats White Paper No. 1.

portal, Pay.gov. In response to demand for shorter *Bulletin* articles on topics of current interest, a couple such articles were published in 2007. In addition, the publications team initiated a pilot project with an IRS field office to image historic issues of the *Bulletin*, published before the birth of Tax Stats, for upload to the web. By early 2008, almost 70 historic issues were uploaded to Tax Stats as searchable PDFs, with other SOI publications scheduled for upload later in 2008. Finally, the cover designs for the *Bulletin,* the *IRS Data Book* and the *IRS Research Bulletin* were updated and given a modern look.

Future Initiatives in Statistical Dissemination

The future of statistical dissemination of SOI products is exciting. Several initiatives are in progress or planned for 2008 and 2009. SOI is developing an application that will allow its customers to create custom tables and basic graphics via the web based on previously tabulated or public-use data.[5] In addition, SOI's webmaster is developing web pages that present simple statistical snapshots of various subject-matter areas for access by customers who are less familiar with tax data. This "layering" project will increase accessibility of SOI data to a broader audience. SOI's plans for dissemination also include development of web pages that target an untapped population of potential SOI customers—children, students and teachers; web pages will include interesting facts on income and taxes suitable for various age and education groups. While documentation or metadata for SOI data sets are available via web in a few subject-matter areas, availability will expand to include more areas over time.

▶ Conclusion

The Statistics of Income, a division of the U.S. Internal Revenue Service and one component in the decentralized U.S. statistical system, is the primary source of data on federal income taxation. With its proud history of disseminating statistical products, which began in 1918 with the first print report on personal and corporate income taxes, SOI has forged ahead into the 21st century. The recent journey from a predominantly print-centric publishing environment, in which web dissemination was an afterthought, to an environment in which print and web dissemination are seen as near equals, has been a challenging one. A measure of SOI's success is the development of one dissemination process that includes both print and web, in most cases.

As SOI looks to the future and all that is possible in terms of statistical dissemination, we keep our eyes on the goal, succinctly stated by the U.S. Office of Management and Budget, Executive Office of the President in April 2008: "(D)ata must be objective and free of bias in their presentation and available to all in forms that are readily accessible and understandable."[6] It is not an easy mandate, but one that we—SOI leadership, subject-matter experts and those directly involved in publication via print and web—will continue to pursue with great excitement.

[5] A modified version of the application will allow SOI staff to tabulate microdata without prior knowledge of statistical software packages, such as SAS and SPSS.

[6] Statistical Policy Directive No. 4: Release and Dissemination of Statistical Products Produced by Federal Statistical Agencies (2008), *Federal Register*, Vol. 73, No. 46.