SOI Studies, Data, and Future Plans
Tom Petska, Director
Statistics of Income (SOI) IRS

“Turning Administration Systems into Information Systems”

SOI Studies, Data and Future Plans

1. Background on SOI in the IRS
2. SOI studies and data availability, focusing on regional data
3. SOI’s website and ways to find SOI data
4. Overview of future plans
Internal Revenue Code of 1986

Section 6108(a)

“The Secretary (i.e., of Treasury) shall prepare and publish not less than annually…statistics available with respect to the operations of the internal revenue laws”

Who Uses SOI Data?

- Treasury’s Office of Tax Analysis (OTA)
- Congressional Joint Committee on Taxation (JCT)
- Commerce’s Bureau of Economic Analysis (BEA) & Census Bureau
- IRS
- Other government agencies
- Academic & private researchers
SOI Budget & Staffing

- Overall budget: $40.5 million
- Total staffing (staff years): 445
  - National Office: 171
    - Economists / Subject Matter: (45%)
    - Computer Specialists: (30%)
    - Math Statisticians: (10%)
    - Other: (15%)
  - Field Sites: 266
    - Editors: (90%)
    - Computer Specialists: (10%)

IRS Returns Processing Pipeline

1040
- Mail Opened
- Returns Sorted
- Checks Deposited
- Code & Edited
- Returns Filed
- Errors Corrected
- Computer Checked
- Computer Entered
- ElectronicallyFiled Returns

Tax Refund Mailed
Return Information Posted At Martinsburg Computing Center
**SOI Service Center Processing**

- Receive lists of SOI-sampled returns from Martinsburg
- Pull returns from files for SOI samples
- Load administrative data for SOI samples into SOI minicomputer network
- Ship sampled returns to SOI processing site
- Match records, key-enter supplemental data, test and resolve discrepancies and errors

**SOI National Office Processing**

- Plan, design and manage processes
- Train field staff in data editing
- Monitor data editing process
- Tabulate, analyze and publish findings
- Weight sample files to national totals
- Error resolve difficult cases
**SOI Statistical Processing**

- **Ogden** Hub Site Programming Edit
- **Fresno** Pull Returns
- **Kansas City** Edit Individual Returns
- **Philadelphia** Pull Returns
- **Andover** Pull Returns
- **Brookhaven**
- **Washington DC** (National Office) SOI Division Headquarters
- **Cincinnati** Hub Site Programming Edit Individual & Estate and Gift Returns
- **Martinsburg** Computing Center
- **SOI Piggybacks” on Tax Compliance Operations**

- IRS mission is to administer the tax laws
- SOI accounts for < 0.4% of IRS $11 billion budget
- SOI is an administrative records system
  - **Pros** -- data are compiled for other purposes
  - **Cons** -- system is not designed for research purposes and is hard to change
**SOI’s Relationship to IRS Tax Compliance Processing**

- Samples drawn from IRS Master Files
- Administrative data for sampled cases copied into SOI’s minicomputer network
- Tax return item content augmented
- Testing and validation are substantially greater
- SOI data are all *pre-audit*

---

**Two Types of “SOI” Data**

- SOI edited sample files
  - *Pros* – rich in content and highly accurate
  - *Cons* – expensive, less timely, and not robust for subnational estimates
- IRS administrative (i.e., population) files
  - *Pros* – less expensive and more timely
  - *Cons* – less content and less accurate
SOI Data Compared to IRS Master File Data

- SOI uses samples; Master Files are populations
- SOI samples select large returns at high rates (SOI avg. size & complexity are greater)
- SOI item content is substantially richer
- SOI testing and validation are much greater

Challenges in Producing Tax Statistics

- Statistics are secondary priority in large administrative agency
- Primary customers are external
- Complex organizational structure affects statistical programs
- Tax law changes frequently
- Financial information reported can be complex
### Population and SOI Sample Sizes

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual (TY 2005)</td>
<td>134,494,440</td>
<td>526,601</td>
</tr>
<tr>
<td>Cross-sectional</td>
<td>292,966</td>
<td></td>
</tr>
<tr>
<td>BEA Supplementary</td>
<td>2,439</td>
<td></td>
</tr>
<tr>
<td>Individual Edit Panel</td>
<td>127,096</td>
<td></td>
</tr>
<tr>
<td>Individual Unedited Panel</td>
<td>421,893</td>
<td></td>
</tr>
<tr>
<td>Corporation (TY 2004)</td>
<td>5,980,702</td>
<td>146,269</td>
</tr>
<tr>
<td>Partnership (TY 2005)</td>
<td>2,949,576</td>
<td>44,241</td>
</tr>
<tr>
<td>Tax-exempt Org. (TY 2004)</td>
<td>390,425</td>
<td>21,739</td>
</tr>
<tr>
<td>Private Foundation (TY 2004)</td>
<td>84,216</td>
<td>11,451</td>
</tr>
<tr>
<td>Unrelated Business (TY 2004)</td>
<td>38,275</td>
<td>7,898</td>
</tr>
<tr>
<td>Estate (TY 2006)</td>
<td>56,696</td>
<td>13,488</td>
</tr>
<tr>
<td>Gift (TY 2006)</td>
<td>264,013</td>
<td>9,217</td>
</tr>
<tr>
<td>Split Interest Trust (TY 2006)</td>
<td>124,632</td>
<td>12,466</td>
</tr>
</tbody>
</table>

### A Typical SOI Sample Design

<table>
<thead>
<tr>
<th>Asset size</th>
<th>Population</th>
<th>Sample</th>
<th>Rate (%)</th>
<th>Weight (1/r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Totals</td>
<td>1,600,000</td>
<td>25,000</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>$10 mil+</td>
<td>10,000</td>
<td>10,000</td>
<td>100.0%</td>
<td>1</td>
</tr>
<tr>
<td>$10m&lt;1m</td>
<td>90,000</td>
<td>9,000</td>
<td>10.0%</td>
<td>10</td>
</tr>
<tr>
<td>$1m&lt;0.1m</td>
<td>500,000</td>
<td>5,000</td>
<td>1.0%</td>
<td>100</td>
</tr>
<tr>
<td>&lt; $0.1m</td>
<td>1,000,000</td>
<td>1,000</td>
<td>0.1%</td>
<td>1,000</td>
</tr>
</tbody>
</table>
**Who Has Access to IRS Microdata?**

- OTA & JCT have full access
- BEA primarily for corporations
- CBO primarily for Social Security analysis
- Census has access to some Master File data
- SSA, for earnings histories
- USDA, to conduct the Agriculture Census
- Everyone, to Individual and Tax Exempt public-use files

**SOI Branches**, **Study Areas, and Publications**

<table>
<thead>
<tr>
<th>Individual</th>
<th>Corporation</th>
<th>Special Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proprietorships</td>
<td>Corporations</td>
<td>International</td>
</tr>
<tr>
<td>Panel Studies</td>
<td>S corporations</td>
<td>Tax-exempts</td>
</tr>
<tr>
<td>High Income/No Tax</td>
<td>Partnerships</td>
<td>Estate and Gift</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Activity Bonds</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Excise Taxes</td>
</tr>
<tr>
<td><strong>Complete Report</strong></td>
<td><strong>Complete Report</strong></td>
<td><strong>SOI Bulletin</strong></td>
</tr>
<tr>
<td><strong>Source Book</strong></td>
<td><strong>Source Book</strong></td>
<td><strong>Research Report</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>IRS Data Book</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Tax Stats Website</strong></td>
</tr>
</tbody>
</table>

(*) In addition to the Statistical Computing Branch
SOI Publications

- SOI Bulletin
- Individual and Corporation Complete Reports
- Corporation Source Book
- Special Studies in Federal Tax Statistics (SOI Research Papers)

Other Publications not based on SOI data

- IRS Data Book
- Research Bulletin (proceedings of the IRS Research Conference)

Other Data Products and Services

- Tax Stats Web Site -- SOI data on the Internet
  - Over 6,600 Files Available - Individual, Corporation, Partnership, Exempt Organizations, and more.
- Statistical Information Services (SIS) Office
- Reimbursable studies
Data by Geographic Area

• Individuals
  – State
  – County
  – Migration Flows
  – ZIP Code
• Tax exempt organizations
• Businesses (forthcoming)
2005 Individual Income Tax Data By State

By Size of AGI including Charitable Contributions, Earned Income Credit, Alternative Minimum Tax, and Income Tax

Source: Spring 2007 SOI Bulletin

2005/2006 Migration Data – State-to-State and County-to-County

- Number of Returns
- Number of Exemptions
- Aggregate Adjusted Gross Income (AGI)

Source: Contact SIS Office
2005 County Income Data

• Number of Exemptions
• Aggregate Adjusted Gross Income
• Wages and Salaries
• Dividends
• Interest Received

– Source: Contact SIS Office

2005 Zip Code Data

• Income and Tax Items, such as --
  – Charitable Contributions and
  – Alternative Minimum Tax
• Number of returns with Schedules C and F
• All by AGI Size Classes

– Source: Contact SIS Office
2006 IRS Data Book

- Data from various offices of IRS
- Available on the Tax Stats Website
- FY 2007 data released by April 2008

Recent Improvements for Customers

- Integrated Business data
- Corporation Source Book multi-year flat files
- Continued improvements to the Tax Stats
- List-serve announcements of new data products and studies
- Payments by Credit Card
## Number of Requests by Type of Organization
### July, 2006 - June, 2007

<table>
<thead>
<tr>
<th>Organization Type</th>
<th>Number of Requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>2,300</td>
</tr>
<tr>
<td>Consultant/Research</td>
<td>483</td>
</tr>
<tr>
<td>Academia</td>
<td>313</td>
</tr>
<tr>
<td>Internal Revenue Service</td>
<td>297</td>
</tr>
<tr>
<td>Media Relations</td>
<td>192</td>
</tr>
<tr>
<td>Private Citizen</td>
<td>247</td>
</tr>
<tr>
<td>State/Local Government</td>
<td>189</td>
</tr>
<tr>
<td>Media (Direct)</td>
<td>181</td>
</tr>
<tr>
<td>Corporation</td>
<td>158</td>
</tr>
<tr>
<td>Federal Government</td>
<td>109</td>
</tr>
<tr>
<td>Nonprofit</td>
<td>99</td>
</tr>
<tr>
<td>Financial Services</td>
<td>96</td>
</tr>
<tr>
<td>Congress</td>
<td>53</td>
</tr>
<tr>
<td>Other</td>
<td>75</td>
</tr>
</tbody>
</table>

## Frontiers in Tax Statistics

- Increased emphasis on longitudinal analysis
- Greater use of master file population data
- Expanded use of e-filed returns
- Expanded web presence
- Less people-intensive data edits
- Expanded Public Use Files
- Synthetic data?
Planning our Future –
SOI’s 2016 Strategic Planning Initiative

1. Customers -- continued expansion
2. Products and services – more data to the public
3. Information processing – leveraging technologies and e-filing
4. Human capital – refining staffing needs

1. Expanding our Customer Base

- Recognition of OTA as JCT as primary customers
- Designation of important national needs of BEA, FRB, and Census
- Formal recognition of IRS, researchers, and others (including the public) as important customers
- Continued expansion beyond current group
- How does this impact resources?
2. **Improving Products and Services**

- New data files (and linkages) for primary customers
- Better, more comprehensive, metadata
- Improvements in public dissemination
  - Enhancements to Website
  - More Public Use microdata
  - Rethinking publications

3. **Leveraging Technologies**

- Greater use of IRS population files
- Leveraging expanded e-filing
  - Scanning residual paper
- Internalizing sampling operations
- Developing smart data edits
4. Developing Human Capital

- Is there a more efficient organizational structure?
- What critical job skills and career paths will we need?
- How to attract, develop, and retain the right people?
- Can we expand the location of work and be more people/family friendly?
- How to recruit and retain the best diverse workforce?

Thanks for your attention!

For questions about SOI data or services—
Contact our SIS Office:
Phone: 202-874-0410
E-mail: sis@irs.gov

Or contact—
Tom Petska, SOI Director
PO Box 2608
Washington, DC 20013-2608
thomas.b.petska@irs.gov