Exploring Researcher Access to U.S. Tax MicroData: Constraints and Potential

Nick Greenia
Statistics of Income Division
U.S. Internal Revenue Service

Note: The views expressed in this presentation are those of the presenter, and may not represent official positions of the Internal Revenue Service.

Workshop on Data Access
August 20-21, 2007

Some Reasons for Researcher Access

- Statistical office's value added is changing – need to be more than a data collector/processor
- Diminishing returns for public use files: expense of protecting data + reduced utility of increasing level of data anonymity
- No Business Data Public Use File in sight
- Complex research questions may require access to microdata (not necessarily with identifiers)

Reasons for Researcher Access (cont.)

- Technological improvements in remote, off-site access
- Data "discrepancies" across different agencies' datasets
- Need for transparency in data-driven policy-making, peer review of analytical results
- Researchers need linked employer/employee microdata, over time

The Importance of Tax Data

- Administration of the tax system (funding government spending)
- Invaluable and economical input to statistical programs
  - Sampling frame
  - Rich financial detail on income, wealth
  - Some demographic data
- Compliance aspect ensures certain level of population coverage, data accuracy
- Voluntary compliance viewed as dependent upon protection – both perceived and actual – of taxpayer confidentiality

Constraints to Tax Data Access

- Statutory Basis → No statute: No Access
- Unlike some statistical data
  - Tax data treated as homogeneous or "equally" sensitive; for example, Person's Name = Business Tax Liability
  - No statute of limitations enabling historical analysis of "safe" tax data
  - Bottom line: all tax data must be protected in perpetuity (even though Sensitive But Unclassified security status)

U.S. Statutes Enabling Statistical Access to Tax Data

1. Outside Statistical Use; e.g., U.S. Census Bureau
2. Tax Analysis (public finance) – Legislative and Executive Branches
3. Tax Administration – but this includes research and statistical functions
   A. Mandated tax statistics
   B. Outside special statistical requests
   C. Contractors for tax administration (including statistical/research activities)

Little Apparent Statutory Flexibility

- Statistical purpose restricted to a handful of agencies, virtually the same after 3 decades
- Most tax analysis [public finance] questions are already tightly assigned to specific offices in the executive and legislative branches
- Data parity access may be jealously guarded monopoly for three reasons:
  1. Control of policy analysis
  2. Competing analyses may be construed as "disruptive" to the legislative process
  3. Confidentiality concerns (see statutory accesses)

Barriers to Statistical MicroData Access

- Statutory – obvious
- Agency Policy – not so obvious
- Researcher Community
  - Resignation, innured to no access?
  - Lack of coordination, representative voice
- Government executive level – institutional disincentives may adversely affect will to innovate
  - Often easier to do nothing perfectly than something imperfectly
  - Mistakes are punished, success seldom rewarded meaningfully
  - Third party scrutiny – media, etc. -- may be career-ending

Dilemma

- Risk takers tend not to be in government, government controls data access, expanding data access entails risk

More Barriers

- Comfort in apparent statutory constraints
  - May provide institutional justification for inertia
- Institutional reluctance in some statistical offices
  - to share data [preserve turf, control processing methodology]
  - to regulate and monitor as part of safeguarding responsibility (an administrative task)
  - to compete/combine programs: some might disappear, be subsumed (maybe offices, too?)
  - for provocative dialogue on microdata access for outsiders vs. aggregate analysis
  - to engage microdata customer community: “Not Invented Here” (NIH) syndrome, fear of criticism.

The “Final” Frontier

- Historical taboo: Policy and Analysis
- Statistical offices tasked to collect survey/census data – not analyze it, right?
- Analysis means policy, no?

Enough Constraints!! Where's the Potential?

- Technology to the rescue: data processing costs going through the floor – at least for the NTO's statistical office.
- What is the future of a statistical office if need for data collector/processor decreasing, but the statute seems so intractable?
- But is statute outpaced by technology?

Enough Constraints -- Where's the Potential?

- "Something unknown is doing we don't know what." – A.S. Eddington

- "The only reason for time is so that everything doesn't happen at once." – A. Einstein

Practice Makes “Perfect” (Or at least Better)

- Narrow statutory interpretation of analysis access cannot keep up with the incessant demand for information by decision makers:
- Past 20+ years: analytical papers by “data processors & collectors”; e.g., the NTO’s statistical office
- Past 10+ years: two additions for statistical access to tax data statute: CBO and NASS
- Past 7 years: Census Research Data Centers
- Past 1 year: NIST/NORC Data Enclave
- → → Controlled expanded statistical and analytical access need not be a zero sum game

Statutory Potential for Research & Statistical Access

* Possible enabling statutes for tax administration research and statistical functions
  - Outside special statistical requests (similar to the recent policy statement promulgated by the U.S. Census Bureau on special tabulations for outside requests, sometimes as unofficial statistics)
  - Contractors for statistical/research activities
* What might be some benefits for statistical tax administration?

Benefits of Outside Access for Statistical Tax Administration

- Analytical value added for the NTO's statistical office – to be relevant for the future, needs to be more than a data processor (like the US Census Bureau).
- Peer review of analytical results may help government avoid perception of "monopolistic myopia", foster transparency in data-driven policy-making
- Nurtures statistical office's professional skills: how to think about not only how to process data, but also different questions that captured data may help answer across forms, data systems, and even agencies.

The Paradox and Limits of Outsider [Research/Statistical] Access

- Q#1: Does unlimited access to confidential data – even if for vetted researchers & even if always protected against unauthorized disclosure -- make confidentiality protection meaningless? Perceptibly?
- Q#2: What is the limit of access that can be managed [by government]?
- Q#3: Is there an alternative process?

Future Outlook

- Cornerstones of confidentiality protection:
  - Safeguard the data: includes authorized purpose of access
  - Some limit on access
- Seek and approve only the “best” research
- Data access: a scarce resource requiring employment of economics allocation process
- More effective coordination needed inter-agency and between agency-researcher communities
- “Outside” researchers and “inside” bureaucrats need to better understand each other’s worlds:
  - Bureaucrats control data access
  - Researchers may be bureaucrats’ future