The Massachusetts Department of Revenue’s Electronic Data Warehouse

Howard Merkowitz & Kazim P. Özyurt

Office of Tax Policy Analysis
Massachusetts Department of Revenue

Background

• Who Developed It: The MA DOR and Revenue Solutions, Inc. (RSI) have partnered for the past four years to build (and later extend) an agency-wide data warehouse.
• Goal originally was to
  - continue to narrow the tax gap (Auditing & Compliance, etc.)
• Later it was extended to
  - improve access to data for analysts to support ad hoc analysis, reporting, estimation, and simulation (Office of Tax Policy Analysis).
Several Years ago, DOR wanted to have a centralized tool to address the non-filing, non-registration, and underreporting compliance gap in a more complete, automated and accurate manner than in the past by

a) identifying non-filed taxpayers using information assembled in the warehouse about their "portfolio" and

b) calculating accurate tax assessments for them.

Steps:

• Before the warehouse project began, a pilot was conducted (to demonstrate the revenue-generating potential and utility of the warehouse)

• After a successful pilot, the project was formally initiated.
Background (continued)

- The warehouse (and the business solutions it supports) has been extended in incremental phases, with each phase reusing the existing infrastructure to deliver value in new areas.

- What began as a warehouse to identify and automatically calculate proposed assessments for individual income taxpayer leads, expanded into a processing platform and data storage for:
  - complex deficiency calculations,
  - performance reports,
  - ad hoc analysis, by using tools such as OLAP technologies
  - compliance case management support

Specific Goals (continued)

- The Data Warehouse uses RSI's Discover-Tax Warehouse as its foundation and uses Discover-Tax Utilities to
  - match data,
  - calculate taxes owed,
  - review portfolios and tax calculations.
Users

• Discovery program examiners responsible for using data to create automated tax calculations for non-filer and deficiency cases

• Information Services analysts who rely on the relational data in the warehouse (refreshed daily) to identify data problems or to respond to query requests far faster than it would take to code custom COBOL programs against the legacy transaction processing system (MASSTAX)

• Statisticians/economists/analysts at the Office of Tax Policy Analysis who have access to raw data and aggregated OLAP cubes in support of analysis questions

Users (continued)

• Audit and Taxpayer Service users who search for taxpayers in the warehouse and may review their full portfolio of available data online

• Auditors who use data mining tools and statistical models to generate their own queries for audit selection

• Auditors who use case management and tracking software to record all details of an audit case in "electronic case folders" supported by the warehouse
Users (continued)

- **Collectors** who use specialized utilities in the warehouse to identify lien and levy sources, offset opportunities, and demographic data.
- **Collections Management** who have implemented a sophisticated risk-based approach to assigning actions to different collections cases by using statistical models and strategy rules that are executed in the warehouse for new cases and which rely on historical taxpayer data in the warehouse.
- **Managers** who use an evolving executive-level reporting dashboard of key measures derived from the warehouse to monitor performance against their goals. *(Still under development.)*

Current Status

- The warehouse has supported tax discovery programs that have generated significant *new revenues* for the Commonwealth - in excess of $325M.

- The warehouse has also begun to "unlock" data that is now *used by more groups* in the agency for reporting, analysis, and compliance management.
  - The requests for additional data have grown substantially.
  - The warehouse itself has *grown to over 5 terabytes* of data and includes several hundred data sources, with additional users coming with *additional needs* for more current and more varied sources of data.
Current Status (Continued)

- The transition to new data sources, analysis tools, and data-driven decision making techniques for various users in the agency has had certain challenges, relating to testing, training, procedures, and communication.

Data

Easily accessible & more frequently available
- *Dor Data* (Return Data, Revenue Accounting Data, Financial Transaction Data, etc., Aggregate collection data by tax type
- *Non-DOR Data* (RMV Registration data, license data, IRS (IRTF, IMF) data)
Data (continued)

- Data delivery used to take up to 2-3 weeks
- Now, it takes 2-3 business days at most, often less.
- The data can be easily queried and transferred to desired platforms.
- Data is refreshed and updated frequently

OLAP

(Online Analytical Process Tool)

- Data Warehouse has OLAP functionality used by Office of Tax Policy Analysis.
- Has drill-down functions for Revenue Accounting data, and others.
**Tax Calculator**
(Under in development)

- Under development for running return data microsimulations

- Could simulate whole or a sub-set of return line to calculate alternative tax liability under various scenarios (exemptions change, deductions change, tax rate change, etc.)

- Work under way for improving its efficiency and usability.

**Future**

- Data warehouse is expanding (more data becoming available over-time). More storage space will likely be needed.
- Demand for data increasing by various divisions
- So, there are plans for investing more in the Data Warehouse
- Ultimate goal is to
  - Support a transition towards a more data-driven environment
  - Have even better performance measurement,
  - Have increased compliance management,
  - Improved and effective decision-making.
Flow Diagram

Warehouse Tables