State Taxes

Growing Number of States Turn To Data Mining to Boost Tax Collections

ATLANTA—Just two weeks after sending out its first wave of collection letters generated by a new computer system, the South Carolina Department of Revenue began receiving payment from delinquent taxpayers who had been flying under the agency's radar.

"We've actually gotten a few checks in already," DOR Director Burnie Maybank told BNA April 21. What made that possible is a $2 million computer that contains various databases and compares notes to produce audit selections, a process called data mining.

South Carolina is just the latest of a growing number of states that have had success with data mining, or are looking to it to boost collections.

Data mining takes different sets of information that a state already has, such as motor vehicle records, employment records, state and federal tax forms, drivers' license numbers, business filing of unemployment insurance, and customs records, and compares them to find likely tax cheats.

"It's sort of all the rage for tax collectors for about two to three years now," Maybank said.

South Carolina will collect an extra $100 million in the first five years the new system is in operation, according to Maybank. Those kinds of numbers, he added, would have been impossible for DOR to post before because the agency traditionally had done data comparison manually.

And with its level of full-time equivalent employees down to 1969 levels until last year, the collection "letter flow" had slowed to nothing, Maybank said. Within six months the new system will be generating 3,000 collection letters a week, he said.

Texas Program Implemented in 1998. Billy Hamilton, deputy comptroller of Texas, told BNA May 2 that the state has taken in about $360 million in extra tax collections since a data mining program was implemented in 1998. Hamilton said Texas expects to collect about $70 million a year with the program over time, "once it settles into a more routine thing."

Hamilton said detection efforts focus on companies that are doing business in Texas but have not properly notified the comptroller's office. He said Texas law prohibits disclosure of companies that have been found to be out of compliance using the program.

"We're very much aware that Exxon and Walmart are here," Hamilton said. "It's usually more obscure companies from out of state that are doing business here and have established nexus with us but haven't bothered to get in line."

In some cases, he added, there are "grey areas" where some companies do not realize they have a tax liability.

Massachusetts Looks at Registered Businesses. Massachusetts has had good success with data mining, or data warehousing, as the process also is called. The state has collected an extra $150 million over the past two years using advanced data matching techniques, according to Tim Connolly, spokesman for the Massachusetts Department of Revenue.

Connolly said April 26 that one data mining effort that has been productive compares the state's list of incorporated businesses to firms that have paid their sales tax on taxable items, or in the case of restaurants, the state's meals tax.

Using data mining, the agency looks to see if incorporated businesses are registered with DOR and are collecting and remitting the proper sales or meals taxes. If an incorporated business is not found on the tax rolls, the system automatically generates a notice of failure to file or intent to assess taxes, he said.

Massachusetts also compares U.S. Customs data on individuals who declare personal property worth more

than $5,000 when they re-enter the United States and then do not pay the state's use tax on that property.

Connolly said that in 2003, the year the program was launched, DOR sent out 114,000 notices for failure to file or notice of intent to assess, a fairly substantial number. But as a direct result of the success of the data mining program, that number rose to 215,000 notices in 2004, he said, the most recent year for which numbers are available.

Washington State Cuts Audit Time. The Washington Department of Revenue has created a data warehouse system to identify and audit unregistered businesses that should be paying taxes, Tremaine Smith, assistant director of the department's audit division, said April 28.

The new system, which cost $1.4 million, allows DOR to better target audit efforts and has increased collections for the agency, Smith said. Data warehousing efforts generated $3.2 million directly and $7 million indirectly in fiscal 2004, its first full year of operation, through identification of unregistered businesses, he said.

The audit division's limited scope audit program formerly focused on some, but not necessarily all, taxes paid by businesses. Data matching was significantly expanded, doubling recovery per hour, Smith said.

For example, in 2004 the agency increased audit recovery by an average of $302 per hour to $708 per hour, up substantially from an average $481 per hour in fiscal years 2000 through 2003. Also last year the average audit time dropped to 19 hours per audit, from 30 hours per audit in 2000-2003, according to Smith.

In addition, the amount of time needed to conduct data analysis queries "went from several months to seconds and eliminated the intervention of other divisions' employees," Smith said.

North Carolina Looking at Data Mining. North Carolina is considering using data mining, an official said. Alan P. Felton, North Carolina's assistant secretary of revenue for tax compliance, said April 22 officials have seen a demonstration of a data mining product and the state is "committed to doing something similar" in the future to allow for a better look at "patterns of noncompliance."

Verenda Smith, a government affairs associate with the Federation of Tax Administrators, an organization of state tax administrators, said that the trend toward data mining "just makes sense.

"Now that computers have gained some capacity and can do some things better than they used to do them, you take a look at it and say how can my system be tweaked to gain this power," she said April 25.

"Data mining is just an enhanced ability to take information you probably already got on hand but actually do something with it," Smith added.

A big plus is that compliant taxpayers can be weeded out of the system, further focusing it on probable noncompliers and underreporters, Smith said.

"The more information you've got the better able you are to target those taxpayers you need to take a closer look at," she said, and then audit.

States always have had access to the information they are cross matching, but have been unable to deal with because of its sheer bulk. "The goal is to use this information in a meaningful way," she said.

It is not easy knowing how to make good use of data mining, Smith said. "It's something the states are putting some thought into and that's why they're talking about it," she said, adding that many are put off by the cost.

Data Matching Goes to New Level. Paul Panariello, vice president of Pembroke, Ma.-based Revenue Solutions Inc., which provides data mining services and products to state revenue agencies, including Massachusetts, New Mexico, and South Carolina, told BNA April 25 that data matching is not new but has gone to a new level.

"The states for years have been doing data matching. It's not a secret what they've been doing," he said, although most of that was much more "single-threaded," taking one data source, such as federal tax returns, matching it to their records, and working the exceptions.

Now, multiple pieces of data can be added together and state tax auditors are better able to ferret out who is paying their fair share, and who is not.

"The more information you bring together about a person or a business the better choices you can make on their compliance," Panariello said. "The goal is to build the pieces of the puzzle together better for them with the data they've always had access to," he added.

"Data warehousing has taken it to a new level. Our matching is more sophisticated and we can do a lot more data parsing," he said.

To ensure privacy, state data warehousing systems are not connected to the Internet, Panariello said. Every touch to the systems they install is monitored and logged, he said.

There are also many off-the-shelf data warehousing tools and some states are simply doing it on their own, he noted.

BY BARNEY TUMEY, ANDREW BALLARD (RALEIGH, N.C.),
KURT FERNANDEZ (AUSTIN, TEXAS),
MARTHA KESSLER (BOSTON),
AND NAN NETHERTON (SEATTLE)