



Data Mining and Data Matching



MSATA 2008

Andrew Kapp (KDOR)

Aron Montaini (KDOR)

James Harrel (KDOR)



Brief

The Kansas Department of Revenue's Business Intelligence Team, in its current state, is four years old. The team has initiated, designed, and administrates several key agency programs, and continues to expand its capabilities and impact on the successful operation of the Department.

This presentation will outline...

- Data we have used
- What we have done
- What we have learned
- What's next



Data

Currently have access to and use the following data sources...

- 27 State Agencies – Professional Licensee
- Tax– Revenue, Labor, SOS Registrations
- Master Death File
- Driver’s License File
- Watercraft, Vehicles, Dealerships
- Businesses – Food Retail, Restaurants, and other regulated business types.
- W2s and 1099s
- State Vendors and Employees



Warehouse



Data storage and update cycles.

- Depending on primary use, data is stored in several different formats:
 - Delimited text files
 - Oracle tables
 - MS Access databases
 - MS Excel worksheets
 - SAS
- Different batch cycles used to refresh data as needed.



Uses

- Compliance matching
 - Licensee
 - W2/1099/DOL match
 - Business to income compare
 - Compensating use registration program
 - Tax Clearance
- Data mining
 - Data purification and verification
- Taxpayer research (MainPage)
- Research
 - Identifying potential data uses





Matching



Licensee:

- 7,445 filers identified
 - 5,232 returns filed
 - \$7.3 Million discovered (+ \$6.7 Mill. Existing)

Business to income:

- 331 filers identified
 - 83 returns filed
 - \$180,000 discovered



Matching



Compensating Use:

- 5,024 businesses identified
 - Single campaign
 - 200 new registrations
 - \$12,000 discovered

W2/1099/DOL:

- 99,000 filers identified (only 6,000 processed)
 - 1,600 returns filed
 - \$1.4 Million Discovered (week 30)



Matching

Tax Clearance:

- 58,000 filers matched (year 5)
 - Continuous matching
 - 3,400 returns filed
 - \$3.8 Million discovered

What is “Tax Clearance”

- Purchasing
- New Employee
- Existing Employee





Research

Follow up research,

- **Tax Clearance**

- **18% Vendors Denied**

- 40% become approved and become non-compliant again within six months.

- **27% New Employees Denied**

- 11% continued non-compliance
 - KDOR maintains 100% compliance.



Research



Follow up research,

- Business to Income

- Data: Business and Individual tax accounts
- High rate of data errors (old system conversions)
- Older businesses responded better to initial
- Correspondence
- Several businesses were closed due to inactivity.



Research

Follow up research,

- W2/1099/DOL

- Data sources W2, 1099, DOL, DL

- Worst Response

- Only DOL Data with high income (100K+)

- 1099 only income

- Average Response

- DOL and one other source

- Best Response

- Three or more sources and prior filings





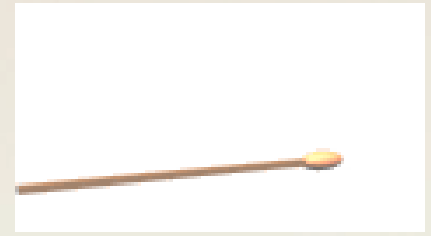
Research

Follow up research,

- Professional Licensee
 - Licensing authorities
 - Worst Response
 - Where data was incomplete
 - Average Response
 - Good data available – frequent renewals
 - Best Response
 - Higher profile – Attorneys, CPAs



Matching



Planned matches:

- Tax Clearance will greatly expand (Web)
 - Grants
 - Private Contracts
 - Professional Licensees
- Businesses
 - Restaurants, Food Sales, Construction
- Profiling
 - Using statistics to identify suspicious scenarios (scoring)
 - Fraud Warehouse (already started)



Questions and Comments



Contacts:

Andrew_kapp@kdor.state.ks.us

Aron_montaini@kdor.state.ks.us

James_harrel@kdor.state.ks.us

