



FTA Conference Using Identity Data to Increase Revenue Collection

Oriane Alves, Business Development
August 6, 2007

© 2007 Identity Systems, a division of Nokia



Company Overview

- **A pioneer in name search and matching**
 - First COTS delivered in 1986
- **A leader in the global market for 20 years**
 - Over 500 clients worldwide
 - Over 60 languages and countries supported
- **Leadership in the Identity Resolution space**
 - Identity Search Server (ISS)
 - Data Clustering Engine (DCE)
 - SSA-NAME3
 - Address Standardization Module (ASM)
- **Continued pace-setter in the industry**
 - 2005 DM Review World Class Solutions Award
 - Uniting Hurricane Katrina victims and families
 - KM World's top "100 Companies that matter in 2006"
- **Identity Systems, Inc. is a Nokia company**

© 2007 Identity Systems, a division of Nokia





Proven Solutions

Corporate



experian



EQUIFAX

VISA



citigroup

nab



ACXIOM



البنك السعودي للاستثمار

ABN-AMRO

ChoicePoint

JPMorganChase

Government



U.S. Postal
Inspection
Service



IRS



FBI



U.S. Dept. of
Homeland
Security



Colorado
Bureau of
Investigation



State of
Texas,
Comptroller



Florida Dept
of Law
Enforcement



Canada
Border
Services
Agency



Australian
Customs

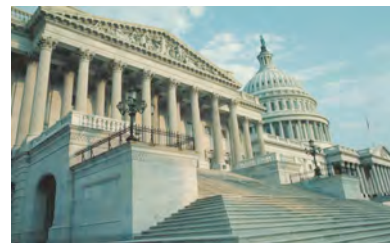


© 2007 Identity Systems, a division of Nokia



State Revenue Department Implementations

- California Franchise Tax Board and Board of Equalization
- District of Columbia
- Georgia Department of Revenue
- Kansas Department of Revenue
- Kansas City, MO Department of Finance
- New York City Department of Finance
- Pennsylvania Department of Revenue
- Ohio Department of Taxation
- Tennessee Dept. of Finance & Administration
- Texas Comptroller of Public Accounts
- Washington Department of Revenue



© 2007 Identity Systems, a division of Nokia



Tax Compliance

- **Government Challenge**
 - Tax Authorities have undiscovered tax revenue from tax evaders or unintentional non-compliance. Data to uncover this revenue spans multiple jurisdictions and sources.
 - Prioritization and minimization of false non-compliance is critical due to limited investigative resources.
- **Underlying Identity Problem**
 - Error and variation exists in tax records
 - Multi-agency data has error and variation due to differing data standards, input mechanisms, data models etc.
 - Error and variation may be deliberate to avoid detection

© 2007 Identity Systems, a division of Nokia



Your Business Requirements

- **Improve match results**
 - Better match from multiple input sources both internal and external
 - Better match of taxpayer information such as SSN/EIN, ID# provided, ID# not provided
 - Enhanced scoring
 - Reduced “false positives”
- **Prevent duplications and consolidate data**
- **Improve accuracy of data**
 - Of data within the data warehouse
 - For compliance purposes to collect outstanding tax liabilities
 - Information received from school districts
- **Enhance the end-user’s experience**

© 2007 Identity Systems, a division of Nokia





Tax Compliance Identity Systems Solution

- **Identity Systems Solution**
 - Identity Systems can provide much more accurate matches within a tax authority's data, and across multiple and disparate sources.
 - This enables tax authorities to identify tax evaders, non-filers, and under-filers.
 - Data mining and warehousing initiatives can benefit from a more accurate view of the taxpayer.
- **Example:**
 - Identity Systems technology dramatically increased revenue discovery at numerous Federal and State tax authorities – and recovered the cost of the project within days of implementation.

© 2007 Identity Systems, a division of Nokia



Existing Techniques – Critical Weaknesses

- **SQL and other DBMS exact searches**
 - Can only find exactly equal or truncated data
- **“Soundex”, “Wild Card” commands**
 - Miss many matches and find many false matches
 - Perform very poorly
- **Query languages and OLAP**
 - Depend on SQL
- **Match codes**
 - Fail when any element is in error
- **Text Retrieval and WEB search engines**
 - Not optimized for identity data

© 2007 Identity Systems, a division of Nokia





What's Wrong with Identity Data?

- Data entry error
- System inadequacies
- Data inconsistency
- Program or data conversion error
- Inherited error from other systems, from external sources

© 2007 Identity Systems, a division of Nokia



Common Errors & Variation

Variation or Error	Example
Sequence errors	•Mark Douglas or Douglas Mark
Involuntary corrections	•Browne – Brown
Concatenated names	•Mary Anne, Maryanne
Nicknames and aliases	•Chris – Christine, Christopher, Tina
Noise	•Full stops, dashes, slashes, titles, apostrophes
Abbreviations	•Wlm/William, Mfg/Manufacturing
Truncations	•Credit Suisse First Bost
Prefix/suffix errors	•MacDonald/McDonald/Donald
Spelling & typing errors	•POrter, Beht

© 2007 Identity Systems, a division of Nokia





Common Errors & Variation

Variation or Error	Example
Transcription mistakes	•Hannah, Hamah
Missing or extra tokens	•George W Smith, George Smith, Smith
Foreign sourced data	•Khader AL Ghamdi, Khadir A. AlGamdey
Unpredictable use of initials	•John Alan Smith, J A Smith
Transposed characters	•Johnson, Jhonson
Localization	•Stanislav Milosovich – Stan Milo
Inaccurate dates	•12/10/1915, 21/10/1951, 10121951, 00001951
Transliteration differences	•Gang, Kang, Kwang
Phonetic errors	•Graeme – Graham

© 2007 Identity Systems, a division of Nokia



English Search Example

SEARCH	William Stuart Harison 117- 2a Jacksen Rd., Easthartford, CT 06987
William Stewart Harison	117 Jackson Road., Hartford, CT 06987 Suite 2A
Doctor Bill Harisen jr	117 Jacson, Room 2a, Hartford East, CT 06984
Harrison William Doctor	2a Jackson Rd, Hartford CT 06987-4573#174978
W. S. Harrison II PhD	Medical Center,117/2A Jackson Rd E. Hartford, NY16987 #17497

© 2007 Identity Systems, a division of Nokia





The 'Real' Numbers Behind Data Quality

Data Characteristics	Organization	# Records	# Duplicates	Duplicates %
Person Name and Address	Marketing system	210,000	48,000	22%
Company Name and Address	Financial Services	113,000	46,000	40%
Person Name, DOB, Address	Government	685,000	21,000	3%
Joint Account Name, Address	Government	150,000	18,000	12%
Person Name and Address	Marketing system	54,000	16,500	30%
Company Name	Government	45,000	9,500	21%
Contact Name & Address	Financial Services	19,000	450	2.50%
Contact Name, Business Name, Address	Financial Services	19,000	220	1.20%
		1,295,000	159,670	12.33% Avg

In a large enough database,
a duplicate rate of 'one-in-a-million'
occurs a *million* times



© 2007 Identity Systems, a division of Nokia



Optimum Solution Criteria

- **“Smart” indexing & key building**
 - Overcome spelling, phonetic, transliteration & multi-country data, missing/out-of-order words and other errors & variation
- **Flexible Search Strategies**
 - Pre-built and customizable search strategies, available at run-time, to balance performance and comprehensiveness of search.
- **Matching algorithms**
 - Emulates a human expert's ability to determine a match based on numerous fields & attributes
- **Speed and Scale**
 - Proven track record of large volume, highly-scalable, critical applications.
 - **Platform independent**
- **Simple to use and deploy**
 - No need for users to learn complex algorithms – standard Population Rules
 - No change to existing databases and tables
 - Complex coding/development not necessary

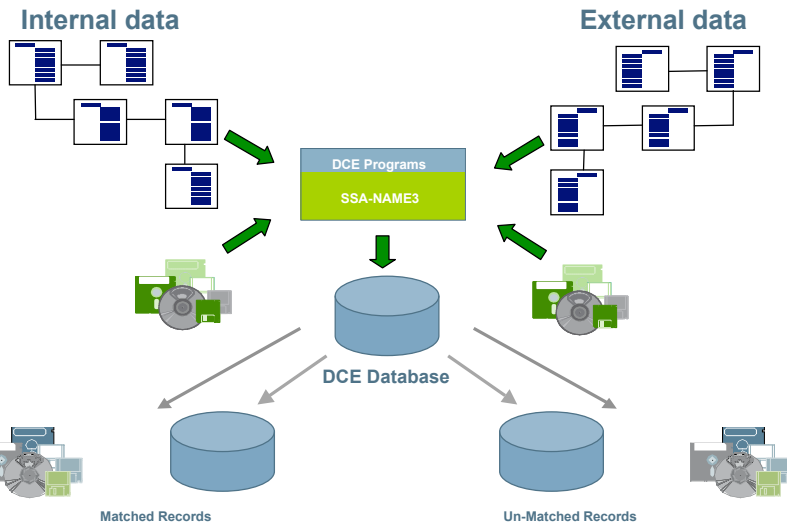
Page 17	ANN Smith B
Ann, Smith Jackson	10MainStSpringvale..9257 5496
Page 119	JACKSON Ann K
Jackson, Ann Smith	10MainStSpringvale..9257 5496
Page 327	SMITH Alan
Smith Ann Jackson	10MainStSpringvale..9257 5496



© 2007 Identity Systems, a division of Nokia



Data Clustering Engine



© 2007 Identity Systems, a division of Nokia



Case Study – Taxation Compliance Office of The Comptroller - Texas



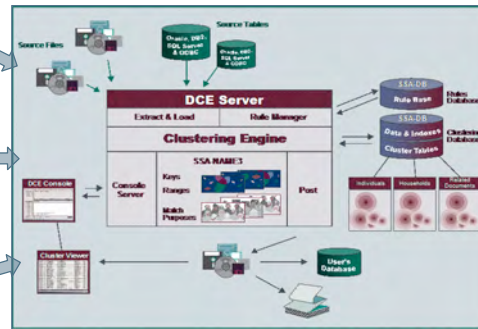
Government Challenge: Texas Comptroller must enforce tax compliance to maximize public revenue, thereby reducing the tax burden on truly compliant taxpayers.

Cross match: IRS Form 1099 for state residents with comptroller tax system
To Investigate: Companies filing 1099 with the IRS as resident of the state but not registered for state franchise tax.

Cross match: FAA airplane registrations with comptroller tax system
To Investigate: FAA registrations with state address not permitted for sales and use tax when the aircraft was purchased.

Cross match: Customs data from Form 7501 with the state tax files
To Investigate: Importers whose final destination is in-state that are not permitted for state sales and use tax; or businesses importing items outside the scope of their sales tax business activities.

SOLUTION



ROI: Texas has collected over \$250 million in additional revenue. They continue to uncover an average of \$1 million additional revenue each week. More accurate identification of non-compliant taxpayers also means less intrusion on truly compliant taxpayers.

© 2007 Identity Systems, a division of Nokia





Case Study Pennsylvania Dept. of Revenue

- Challenge:
 - Need to validate external data against internal data to make it useful
 - External data is provided in a completely different manner of identification
 - Previously done manually, taking too much time with a high rate of failure
 - Timing is critical! It is harder to collect from older data.
- Gain:
 - Using name and address
 - linking of internal records with external records like rental rebate programs, cigarette tax programs, etc.
 - automated extracting and assigning of EIN/SSN to external records to run against internal data
 - Able to take external data and make it useful
 - Doing it in a timely manner
- ROI
 - \$4 million in collections has already been identified

© 2007 Identity Systems, a division of Nokia



Case Study Washington State Dept. of Revenue

- Challenge:
 - Data comparisons and searches were done manually and very inefficiently across multiple systems.
 - The Unregistered Business Study identified 99,000 businesses registered with the IRS not registered with the state. Around \$225 million annually in missed taxes.
 - No tools to "Soft Match" 3rd party data to WA-DOR data
 - Any tax discovery process was slow and expensive.
- Gain:
 - The "Soft Match" capabilities allow better identification of unregistered taxpayers. Ability to compare what an entity is reporting to a 3rd party (IRS or ESD) vs what is being reported to DOR.
 - Can get more valuable matches across the data sets helping identify under or non-payers.
- ROI
 - 10 to 1 ROI during first year: original investment of approx \$1.3 million resulted in \$13.7 million in recoveries.
 - are experiencing a higher per hour return on audit efforts per GMAP, average recovery increased by \$302 per hour
- Future Enhancements
 - Add additional data sources, such as the Yellow Pages. There is potentially an additional \$150 million missing tax risk from businesses that should be, but are not, registered with any state or federal tax administration agency

© 2007 Identity Systems, a division of Nokia





Questions



© 2007 Identity Systems, a division of Nokia



Fran Vogel; Identity Systems - Sales Manager
Oriane Alves; Identity Systems – Business Development

© 2007 Identity Systems, a division of Nokia