360 degree Profiling

-- Using Data Mining to convert information to actionable intelligence

G T Venkateshwar Rao IRS
The message by other tax administrations to improve voluntary compliance

"I Saw What You Did and I Know Who You Are"
Often the tax investigators get sketchy information about
- Some name and address.
- Some number linked to the tax payer like PAN, cell number, vehicle number, Passport no, Aadhar no
- Information on some High value financial transaction like
  - date
  - amount

These bits and pieces need to be developed into actionable intelligence.
Income Tax department in India has large Internal databases -

- Identity particulars - PAN
- Tax payment particulars - OLTAS
- Tax Deduction particulars - TDS
- Returned / assessed incomes - AST
- Particulars of transactions in shares – STT

Large External financial transaction databases

- Telephones
- Property sale/purchase
- Bank information with large cash transaction and Fixed deposits
- Purchase of costly four wheelers
- Spending through credit card info etc
- Spending on travel
- Large insurance premium
- Others
Challenges in processing 3Vs (Variety, Volume, Velocity)

1. No single unique Identifier across all data sources (absence of Citizen ID)

2. Forced to use alternate identifier. The only other alternate identifier is name & address

3. No defined standards for writing name/address
   ★ Names and addresses are subject to variations and transcription errors

4. Large data volumes (multiple data bases of the order 2 to 5 crore each)

5. Data Velocity is very high

6. Previous attempts on processing on name & address were not successful
High Level Process of ITDMS

**INPUT**
ETL PORTION.

**DATA CAPTURE**
COMBINED DATA FROM VARIOUS SOURCES

**PROCESSING**

**DATA VIEW**

**OUTPUT**
ANALYSIS

**DATA QUERY**
FAMILY TREE GENERATION

**SEARCH PORTION**

**Pictorial Report**
AND ANALYSIS

**ANALYSIS**

What data to search

Internal:
- PAN
- AST
- OLTAS

External:
- Credit Card
- Property Sale and Purchases
- Vehicle Purchases
- Passport
- Mobile
- Travel
- Aadhar
Search attributes of an Entity

**Attribute**
- Name
  - Name
  - Fathers name
  - Aliases

**Unique No**
- PAN
- Phone number
- Bank Account
- Passport number
- Aadhar number
- Email
- Vehicle Regn no

**Address**
- Address1
- Address 2
- Address 3
- City
- Locality
- Street name
- Road name

**Others**
- Amount
- Date
- Date of birth
What parameters to search

- Unique No.
- PAN No.
- Vehicle No.
- Aadhar No.
- Bank Account No.
- Date of Birth
- Date of Incorporation

Level and Complexity of Search

- Unique Identifiers
- Combination of Non Unique Identifiers
- Only Non Unique Identifiers
- Vaguely Unique
  - Name Alone
  - Address Alone
  - Etc.
- Reasonably Unique
  - Name + Address
  - Name + Date of Birth
  - Name + Father’s Name Etc.
Data Variety (In name, date of birth, address)

Property

Name: S R Tendulkar
DOB 10/12/1973
Address: 12/123 Javeri Road, Bombay, India
Phone
Email

PAN

Name: S Ramesh Tendulkar
DOB 12/10/1972
Address: 12 Javeri Road, Bombay, India
Phone
Email

Foreign travel

Name: Sachin Tendulkar
DOB 12/10/1973
Address: 12-10-123 Javeri Road, Mumbai, India
Phone
Email

Phone

Name: Sachin R T
DOB 12/11/1973
Address: 5-10 Javeri Road, Mumbai, India
Phone
Email
360° Profile of the tax payer

**Internal Sources**
- PAN
- AST
- PLTAS

**External Sources**
- Phone no
- Passport no
- Driving License no
- Aadhar

**Data Points**
- Names, Alias Names, Organization name
- Address
  - House no
  - Locality
  - City, State, Pincode
- Father Name
- PAN
- Phone no
- Passport no
- Driving License no
- Aadhar

**Combined Data**
- Identity Resolution (IR Engine)

**Entity Resolution supporting combination Of Matching Rules**

**Relationship Resolution**
- Child1
- Child2
- Father
- Sibling1
- Sibling2
- Spouse
- Household entity1
- Household entity1

**Single View of the entity**

**Family members**
- Child1
- Child2
- Father
- Sibling1
- Sibling2
- Spouse
- Household entity1
- Household entity1
Adoption within the department

→ **ITDMS is installed** in all 20 Directorates of Investigation across the country in 2008. Undergone major up gradation increasing the capacity from about 2 Cr to about 10 Cr per location.

→ **ITDMS** has now become-
  - a potent tool for identifying cases for large tax evasion for further investigation
  - part of standard procedure of investigation of tax evasion complaints and pre-search enquiries
ITDMS is handling about 1100 million records and is probably the largest data mining in the country and one of the largest in the world using non unique id like name and address.

It is a quantum leap for non intrusive investigation for detecting tax evasion and helps to spread the message that Indian Tax Administration also knows who you are and what you did.
A complete process reengineering

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability to use approximate/alternate identifier</td>
<td>Limited</td>
<td>Comprehensive</td>
</tr>
<tr>
<td>Grouping of transactions of an entity</td>
<td>Non-existent</td>
<td>Comprehensive</td>
</tr>
<tr>
<td>To know all the entities related to each other</td>
<td>Non-existent</td>
<td>Comprehensive</td>
</tr>
<tr>
<td>Time for the profiling</td>
<td>2 to 3 weeks</td>
<td>Less than 1 hour</td>
</tr>
<tr>
<td>Ability handle large data volumes</td>
<td>Could not handle</td>
<td>Handles With ease</td>
</tr>
<tr>
<td>Ability to intelligently mine data</td>
<td>Not available</td>
<td>Fully capable</td>
</tr>
</tbody>
</table>
Ration-cards (Duplicate)

Demographic data: Name, Father Name, Age, Address

Match on Combination of Head and Family members demo graphic data with and without address
Aadhar Based solution cannot solve all

It is understood that these are proposed to be solved through seeding of Aadhar number.

Aadhar seeding based solution cannot solve the above three (bogus /Duplicate/ ineligible) but can solve some of them.

An efficient Entity Resolution Engine based solution is required in addition to using Aadhar number.
<table>
<thead>
<tr>
<th>CARD_NO</th>
<th>CARDNAME</th>
<th>AGE</th>
<th>ADDRESS</th>
<th>MEMBER_TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>WAP159100100099</td>
<td>Bode Sundar</td>
<td>36</td>
<td>1-5-144/51C INDIRA NAGAR</td>
<td>HEAD</td>
</tr>
<tr>
<td>WAP159100100099</td>
<td>Bode Vinitha</td>
<td>12</td>
<td>1-5-144/51C INDIRA NAGAR</td>
<td>MEMBER</td>
</tr>
<tr>
<td>WAP159100100099</td>
<td>Bode Vishal</td>
<td>15</td>
<td>1-5-144/51C INDIRA NAGAR</td>
<td>MEMBER</td>
</tr>
<tr>
<td>WAP159100100099</td>
<td>Bode Nagamma</td>
<td>28</td>
<td>1-5-144/51C INDIRA NAGAR</td>
<td>MEMBER</td>
</tr>
<tr>
<td>YAP152300600196</td>
<td>Bode Nagamma</td>
<td>32</td>
<td>2-63</td>
<td>HEAD</td>
</tr>
<tr>
<td>YAP152300600196</td>
<td>Bode Vineetha</td>
<td>13</td>
<td>2-63</td>
<td>MEMBER</td>
</tr>
<tr>
<td>YAP152300600196</td>
<td>Bode Vishar</td>
<td>16</td>
<td>2-63</td>
<td>MEMBER</td>
</tr>
<tr>
<td>YAP152300600196</td>
<td>Bode Sundar</td>
<td>36</td>
<td>2-63</td>
<td>MEMBER</td>
</tr>
<tr>
<td>WAP1508032A0246</td>
<td>Dappu Manjula</td>
<td>24</td>
<td>4-112/1 ----</td>
<td>HEAD</td>
</tr>
<tr>
<td>WAP1508032A0246</td>
<td>Dappu Pavanteja</td>
<td>1</td>
<td>4-112/1 ----</td>
<td>MEMBER</td>
</tr>
<tr>
<td>WAP1508032A0246</td>
<td>Dappu Somyasri</td>
<td>2</td>
<td>4-112/1 ----</td>
<td>MEMBER</td>
</tr>
<tr>
<td>WAP1508032A0246</td>
<td>Dappu Kunalkumar</td>
<td>4</td>
<td>4-112/1 ----</td>
<td>MEMBER</td>
</tr>
<tr>
<td>WAP1508032A0246</td>
<td>Dappu Mahender</td>
<td>28</td>
<td>4-112/1 ----</td>
<td>MEMBER</td>
</tr>
<tr>
<td>WAP1588106B0479</td>
<td>Dappu Mahender</td>
<td>29</td>
<td>6-91/1 HARIJANBASTI</td>
<td>HEAD</td>
</tr>
<tr>
<td>WAP1588106B0479</td>
<td>Dappu Pavantej</td>
<td>1</td>
<td>6-91/1 HARIJANBASTI</td>
<td>MEMBER</td>
</tr>
<tr>
<td>WAP1588106B0479</td>
<td>Dappu SOWMYA SREE</td>
<td>2</td>
<td>6-91/1 HARIJANBASTI</td>
<td>MEMBER</td>
</tr>
<tr>
<td>WAP1588106B0479</td>
<td>Dappu Kunal Kumar</td>
<td>3</td>
<td>6-91/1 HARIJANBASTI</td>
<td>MEMBER</td>
</tr>
<tr>
<td>WAP1588106B0479</td>
<td>Dappu Manjula</td>
<td>24</td>
<td>6-91/1 HARIJANBASTI</td>
<td>MEMBER</td>
</tr>
<tr>
<td>WAP1514015A0584</td>
<td>MADHAGONI KRISHNAIAH</td>
<td>36</td>
<td>75 Turkayamjal</td>
<td>HEAD</td>
</tr>
<tr>
<td>WAP1514015A0584</td>
<td>MADHAGONI NAVYA</td>
<td>10</td>
<td>75 Turkayamjal</td>
<td>MEMBER</td>
</tr>
<tr>
<td>WAP1514015A0584</td>
<td>MADHAGONI ANIL</td>
<td>13</td>
<td>75 Turkayamjal</td>
<td>MEMBER</td>
</tr>
<tr>
<td>WAP1514015A0584</td>
<td>MADHAGONI ANUSHA</td>
<td>14</td>
<td>75 Turkayamjal</td>
<td>MEMBER</td>
</tr>
<tr>
<td>WAP1514015A0584</td>
<td>MADHAGONI MANAMMA</td>
<td>30</td>
<td>75 Turkayamjal</td>
<td>MEMBER</td>
</tr>
<tr>
<td>WAP1515162D0070</td>
<td>Madagoni Krishna</td>
<td>32</td>
<td>8-184 LAXMI NAGAR COLONY</td>
<td>HEAD</td>
</tr>
<tr>
<td>WAP1515162D0070</td>
<td>Madagoni Navya</td>
<td>7</td>
<td>8-184 LAXMI NAGAR COLONY</td>
<td>MEMBER</td>
</tr>
<tr>
<td>WAP1515162D0070</td>
<td>Madagoni Anil</td>
<td>9</td>
<td>8-184 LAXMI NAGAR COLONY</td>
<td>MEMBER</td>
</tr>
</tbody>
</table>
Some states have set up SRDH but its utility is not fully exploited.

SRDH utility can be improved substantially for providing 360 Degree view of every citizen with complete exposure about every welfare programme being received in addition the details of employment, family members, Vehicle information, House property etc can be captured which is useful for a variety of purposes including enhancing the tax collections from property tax.

Integrated Household Survey done by Telangana state
యుగానా, నిర్మాణం: సహాయాన్ని లెండరు తెలియజేస్తాడి. మానవ పరిస్థితి మరియు వాయువు సంపాదన పరిస్థితిలో మనకు వెలుగులు లేదా విప్పుల కొరకు పరిపాలన అవుతుంది. దీనితే నాటికి నివాసానికి మరియు పరిస్థితి మాత్రమే సాధ్యం. మనం కోడ్లో చేసిన ఎమ్మెల్యే పరిస్థితి కోసం మన పాత్రలను ఉపయోగిస్తుంది. ఈ పరిస్థితి నాటికి నివాసానికి మరియు పరిస్థితి మాత్రమే సాధ్యం.

అంటే, మాత్రమే: కొన్ని సమయంలో ఆసక్తి, మారుగనికి ఆహారం వాయువు విస్తరణ కాగం ఉంది. మారుగనికి ఆహారం వాయువు విస్తరణ కాగం ఉంది. మనం కోడ్లో చేసిన ఎమ్మెల్యే పరిస్థితి కోసం మన పాత్రలను ఉపయోగిస్తుంది. ఈ పరిస్థితి నాటికి నివాసానికి మరియు పరిస్థితి మాత్రమే సాధ్యం.
Relevance to other intelligence agencies like IB/NIA

- Passport
- PAN
- Mobile no.
- International travel
- Negative List
- Bank A/c info.
Integrated Information Search for Police (MP Police)

Digital Information at PHQ and all stations

Data Mining
- Mobile phone data
- Passport data
- Voter ID
- Aadhar

Text Mining
- Audio
- Video files

Text Mining
- E mails

Text Mining
- FIRs, Case diaries, and all other documents in Word, Excel, Pdf, Ppt

English

Telugu
I-T data mining software gets PM's award

NEW DELHI: S S Khan, Milap Jain and G T V Rao, who developed the Integrated Tax Payer Data Management System (ITDMS), a data-base mining software that provides actionable intelligence to help the Income Tax Department's Investigation Division follow potential violators, got the Prime Minister's Award for excellence in governance and administration as part of Civil Services Day on Wednesday.

The work on the software was started by Khan, the then DG I-T (Investigation), Delhi in 2006. After pilot testing, it was developed as Enterprise Level Solution and rolled out at all 14 stations of DG I-T (investigations) across the country in 2007. In fact, the I-T Department conducted about half a dozen searches based on the actionable intelligence provided by the ITDMS last year.
“With the ITDMS deployed at all the DGsIT, it is expected to improve the data mining and non-intrusive investigative capabilities of the department substantially, Income Tax department has taken head start and is the first enforcement agency in the country to implement a state of art profiling system using sophisticated name search engine on Indian Names."

Shri S S Khan, Member, CBDT
Thank you