



India's Integrated Taxpayer Data Management System (ITDMS)

-A data mining tool for non-intrusive anti-tax evasion work

**Winner of
Prime Minister Award For Excellence
In Public Administration
April 2010**



ITDMS

The presentation is prepared based on the information in public domain
such as in media as well as information circulated during the
Indian Civil Services Day in April 2010

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 (IT-DMS), 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. **ENS**

ITDMS

Successful and Non-Intrusive

- ITDMS is a data mining solution conceived and implemented by the Income Tax Department in India
 - to enable 360 degree view of any entity from the available data with the department primarily to detect potential cases of large tax evasion.
 - To act as a tool for providing actionable intelligence to the investigation units.
 - To promote voluntary compliance.

Evolution of Data bases (1)

- Consequent to the implementation of comprehensive computerisation, several **large internal databases** have come into existence.
 - **Identity particulars-** PAN
 - name, father name, date of birth, address,
 - nature of business, constitution of firm/ company
 - **Tax payment particulars** - OLTAS (On Line Tax Accounting System) from Banks
 - **Tax Deduction particulars** -TDS (from Deductors viz., Employers/ Payers etc)
 - **Returned / assessed incomes-** AST (From taxpayers)
 - **Particulars of transactions in shares** - STT (Security Transaction Tax)
- Other **External databases** like—
 - Telephone directories,
 - MCA databases, -Companies, Directors, Quarterly/Annual Accounts of Companies
 - High Net Worth Individuals


Evolution of Data bases (2)

- **Annual Information Returns (AIR)** introduced in 2005 requiring **third parties** e.g.
 - Banks,
 - Stock exchanges,
 - RBI Bonds
 - Mutual Funds, IPOs
 - Credit card companies,
 - Registrars of Immoveable Propertiesto electronically furnish data of **high value financial transactions entered with them**
- **Challenge** was to-
 - Co-relate the huge information of financial affairs with returned incomes/ tax payments. and
 - **Convert information into actionable intelligence**

Use of data before ITDMS

- Technology to mine huge volume of data was not available.
- The search within the data was ad-hoc in nature and was person specific.
- Full data was not being utilized.
- Since the internal data was not being utilized fully, the department was depending on informers, Tax Evasion Petitions from public etc. These practices were invasive and open to misuse.

Challenges in using data(1)

- **No unique identifier** (citizen ID) across all data sources
 - Though PAN is required to be quoted as unique ID, it is often not quoted, or incorrectly quoted where the transaction is not meant to be accounted
 - **No defined standards for writing name/ address**
 - Names & addresses are subject to variations and transcription errors
 - Poor and uneven data quality
 - **Unstructured external databases** – e.g. in some Name is in one field while in others in three fields
 - **Large volumes** (multiple data bases of 20-50 Million (2-5 crore) records each)
 - **Large false positives** in existing solutions
- 

Challenges in using data(2)

- Information from all possible data sources-internal and external, has to be used on 'as is where is basis' as it was not possible / advisable to go back to the originators of data for correction
- It was felt that unless the issue of matching available data on alternate/ approximate parameters other than PAN is solved valuable data running into millions of records would remain un-utilised

Errors and variations in data - names

Sl. No.	Reason	Examples
1	Spelling	G Venu Gopal Rao / G Veni Gopala Rau
2	Abbreviation	G.V.G.Rao / G.V.Gopal Rao
3	Prefix	Mr. G.Venu gopal Rao Sri. Sree
4	Sequence variation	G Venu Gopal Rao / Venu Gopal Rao G
5	Missing/Extra parts	G Venu Gopal / G venu Rao /G.Gopal rao
6	Splitting	G Venu gopal Rao/ G Venu gopalrao

Even if the information in the database is deemed to be “Perfect” , the name to be searched has to be searched on the basis of information **available**, both of which are bound to be different

Errors and variations in data - addresses

- Large variations in addresses

(for example, all the following refer to the same address)

- PLOT NO 1310 RD NO 65 J HILLS CO OP SOCIETY HYDBAD
 - 1310 ROAD NO 65 JUBLEE HILLS H'BAD
 - RD-65 JUBLEEHILLS, 1310 HYDERABAAD
 - P NO 1310, RD NO65, JUBILEE HILLS, HYDRABAD
- Variations and errors in date field (for example dob-08/03/09)
 - Interchange of day and month– 03/08/09 (dd/mm/yy OR mm/dd/yy)
 - Error in either day or month or year – 18/03/09, 08/08/09, 08/03/90

Prototype of the data mining solution

- Multi skilled **in-house team** set up in Delhi investigation unit
- Team carried out-
 - Requirement analysis
 - Study of search technologies available in market
 - Study of international best practices on profiling
- A prototype was built using a name search engine developed in house for matching names and addresses where PAN was not given or was wrong
- The prototype was
 - tested and deployed at Delhi
 - trial run carried out by the team on Delhi data for one year
 - Proto type customized based on the results and other inputs
 - named Integrated Tax Payer Data Management System (ITDMS)

What is ITDMS

Integrated Tax Payer Data Management System (ITDMS) as a concept refers to

- Generating 360 degree transaction profile of an entity by grouping the information of that entity from all data sources.
- If the entity is individual, wherever possible, build the family tree of the person to include parents, siblings and kin
- To know about all the related entities in which the person being investigated has interest either as a director or partner etc.

Decision for a full fledged solution(1)

- Due to very good results after the implementation of prototype it was decided to go in for a full fledged solution.
- As the software solution was technology intensive and is dependent on availability of powerful name and address search technology, evaluation criteria was decided as Quality & Cost Based System (QCBS).
- Some bidders were short listed based on previous implementation experience of a similar work and similar volume in India and availability of proprietary name search engine for Indian names

Decision for a full fledged solution(2)

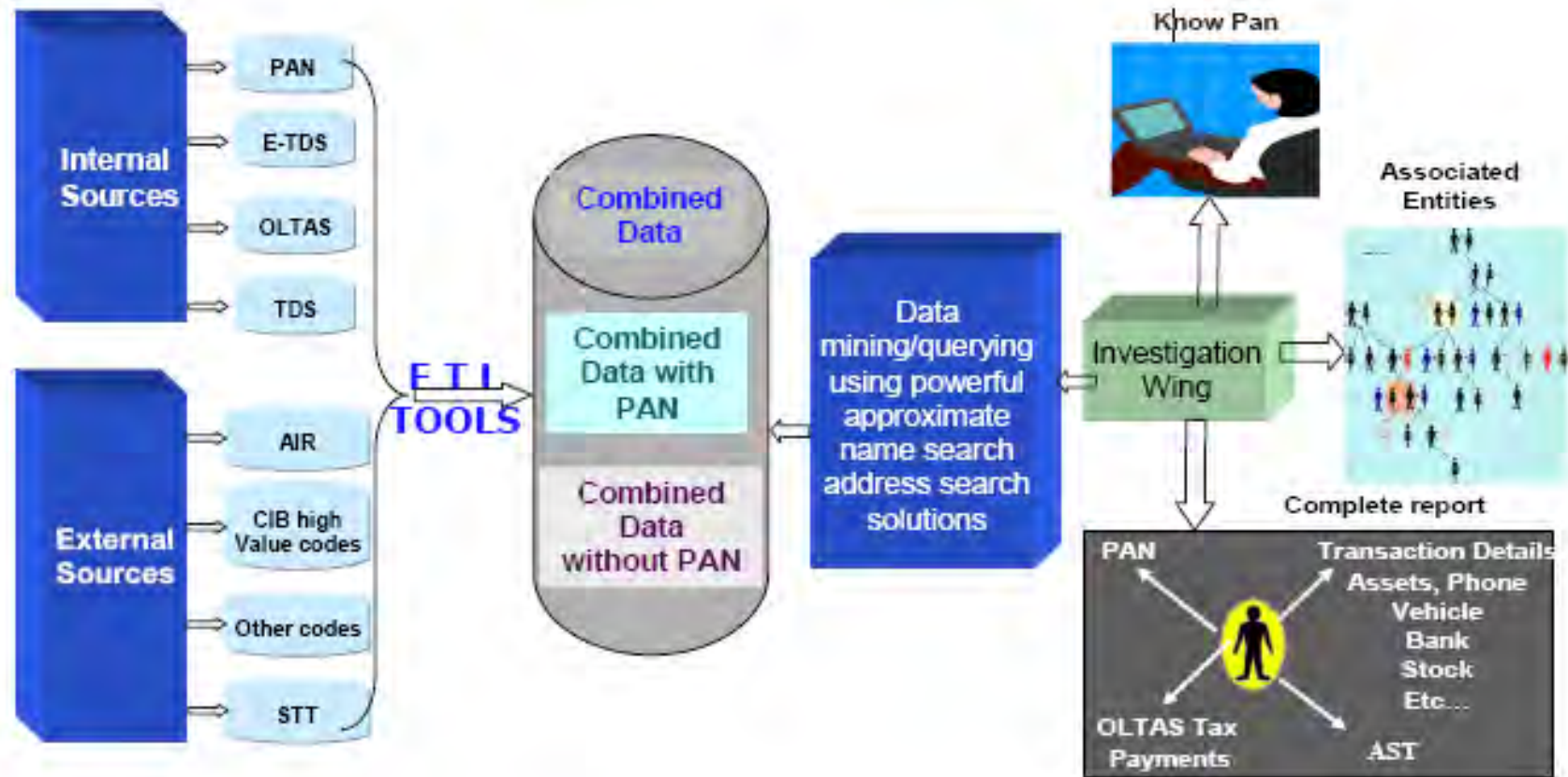
- The short listed bidders did a detailed Proof Of Concept (POC) on the data of 4 million records. The data and the hardware were provided by the department.
- The Proof Of Concept covered 12 different functionalities of the software out of which minimum score was prescribed for four important parameters.
- The project was awarded to the vendor with the best score.

Solution features

The final solution has the following features.

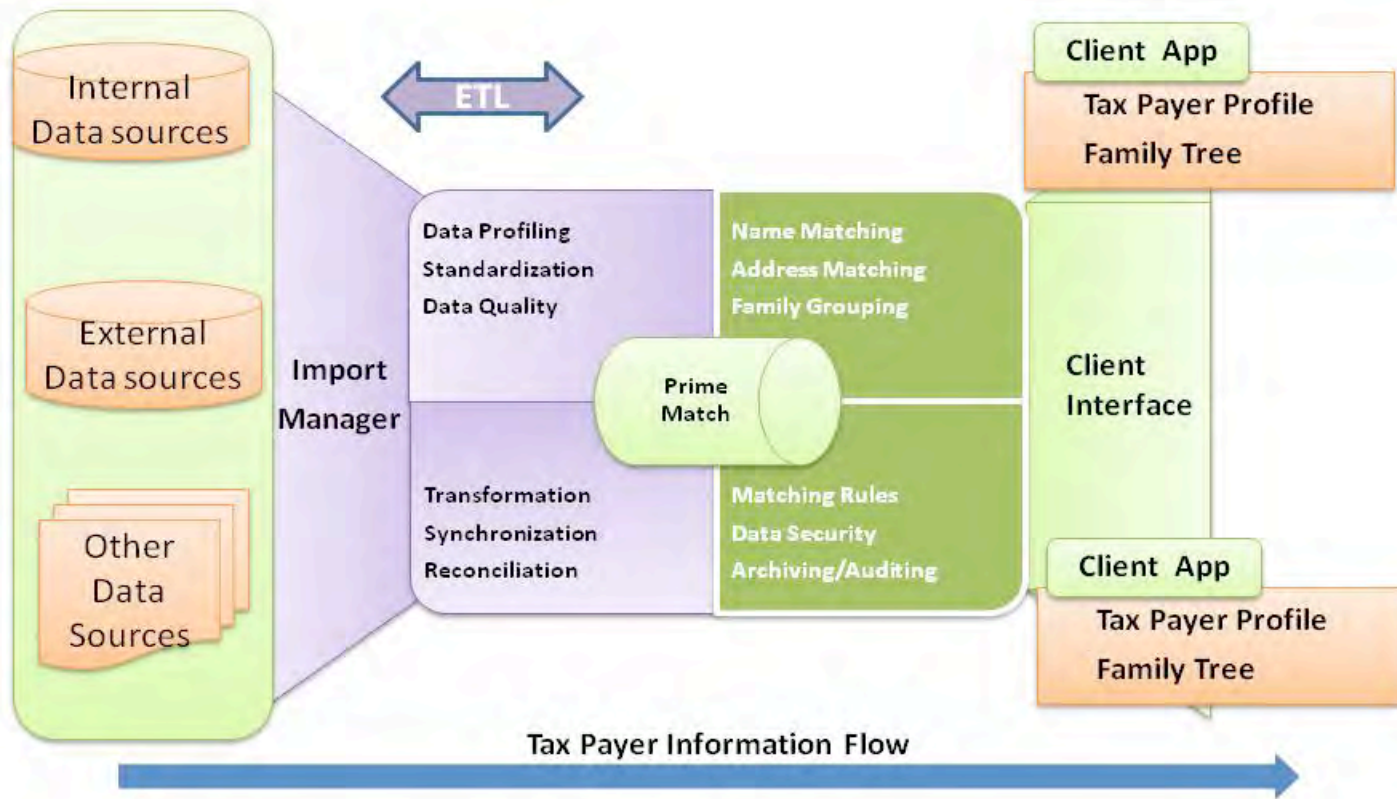
- An inbuilt ETL tool
- An excellent identity search tool , PrimeMatch™ that can handle possible variations in identification parameters such as Name, Address, DOB while dealing with huge data volumes.
- Builds family tree by relating the names and addresses.
- Ranking of the results so that the best results displayed at the top.
- Response time in milliseconds on 20 million records with high recall and precision
- Though the software is OS and database independent, it was implemented on the following as per the requirements of the department.
 - Oracle as the back end database
 - Using Microsoft .Net as a client and J2EE as server platform
 - Dual CPU quad core with 8GB RAM at each location

Solution Architecture – an overview

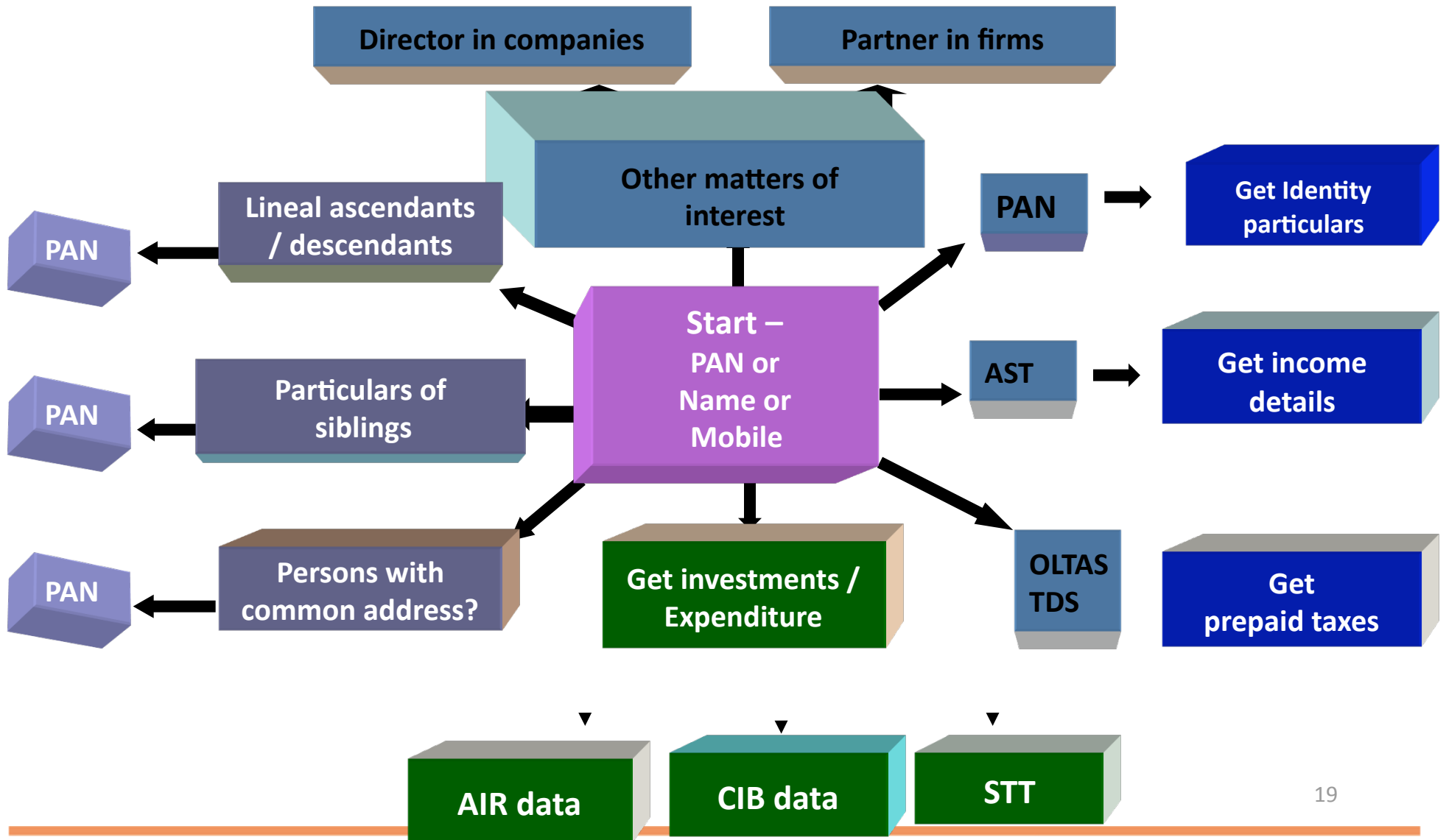




ITDMS --Tax Payer Information Flow



Schematic diagram of ITDMS

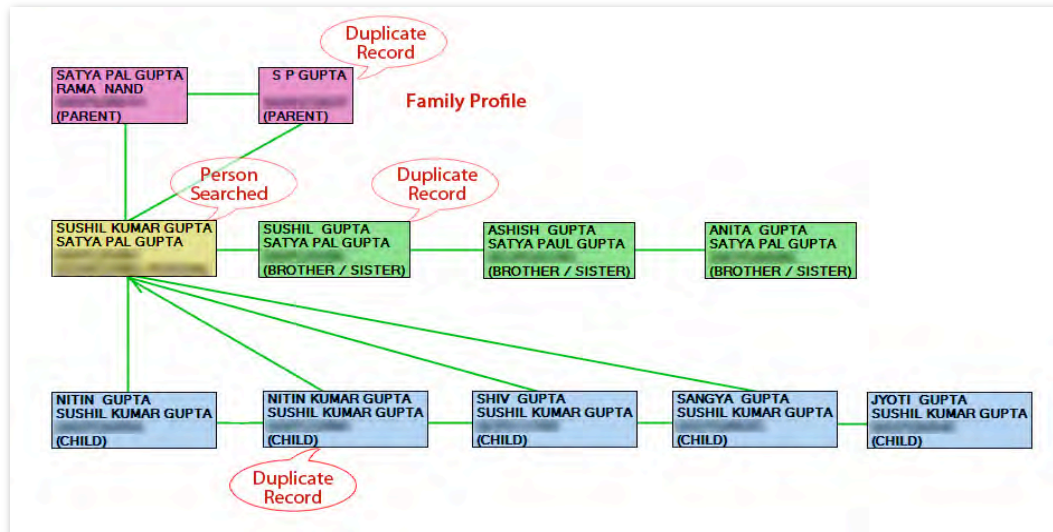


Features and functionalities of ITDMS

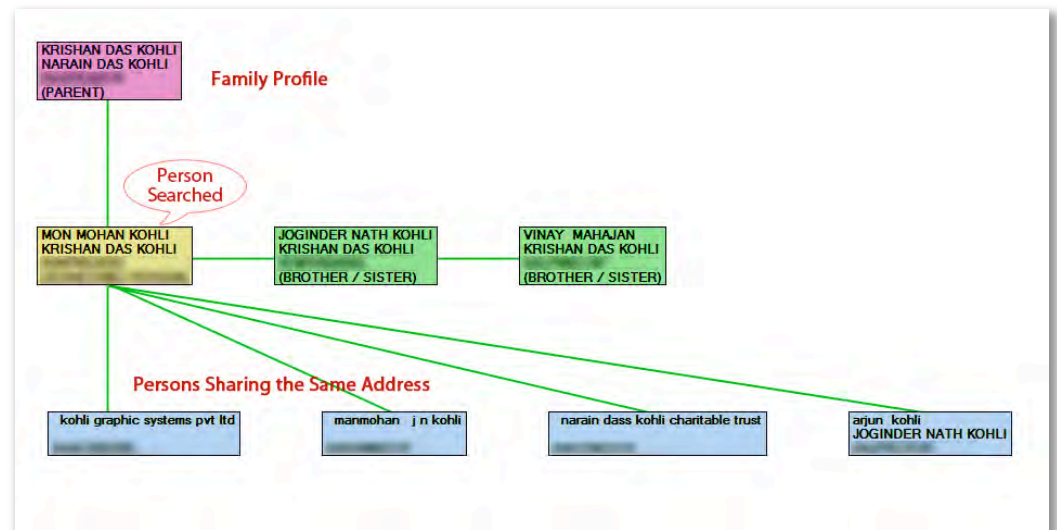
- **Search can start simply from-**
 - a PAN or telephone number or name or part name + address
- **Intelligently searches large databases** of the order of 50 Million(5 crore) in different formats in seconds
 - Using phonetic matching as well as pattern recognition techniques
 - goes behind the data of an entity to search transactions with no PAN/ wrong PAN/ name variations etc identifying suspicious transactions
 - **ranks results** according to the probability of matching
 - Maintains **audit trail** -for internal accountability
- **Creates an integrated profile of taxpayer** showing-
 - Family tree, connected concerns
 - Financial transactions in AIR/STT etc;
 - Incomes disclosed in returns; Taxes paid / deducted at source
- **Re-engineering** of the processes relating to data handling in Investigation Wing

Sample screen shots of family profiles

Sample 1



Sample 2



A complete process reengineering



Parameter



Before



After

Ability to use approximate/alternate identifier	Limited	Comprehensive
Grouping of transactions of an entity	Non-existent	Comprehensive
To know all the entities related to each other	Non-existent	Comprehensive
Time for the profiling	2 to 3 weeks	Less than 1 hour
Ability handle large data volumes	Could not handle	Handles With ease
Ability to intelligently mine data	Not available	Fully capable

Adoption within the department

- **ITDMS is installed** in all 20 Directorates of Investigation across the country
- Development of **User manuals, training and** hand-holding support etc
- **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
 - Tool for information on demand to the officers.
 - To identify the cases for investigation by choice rather than by chance.

Case Studies

- Air Intelligence unit had the mobile phone number and name of person who boarded a flight and suspected to have been carrying cash. Using the ITDMS, it was possible to find out all the entities in which he has interest. By the time he landed, all his business/residential premises were searched resulting in detection of large income on which taxes were not paid.
- In a case of a group dealing in steel industry pre-search profiling through ITDMS brought out new information of connected persons and premises. Subsequent searches resulted in admission of huge undisclosed income of over Rs 500 crore.(\$ 100 million)

Case Studies

- In another case of a Tax evasion petition profiling on ITDMS showed high investments. A letter was issued following which assessee admitted additional income of Rs 80 Crore (\$ 16million).
- A person faked identity and managed to get three unique PANs (Tax Accounting No.) allotted. The ITDMS software could establish the links through family search feature and the connections with a large real estate firm was established. The authorities detected tax evasion of Rs.100 crores.(\$ 20 million).
- Rs.40 Crore(\$ 10 million) was recovered from a stock broker on investigation after the ITDMS showed large demat transactions

Probably the largest data mining in Government

- With the installation of ITDMS all 20 Directorates of Investigation across the country, the total data volume handled is about 300 million (and is growing) using identity resolution name and address.
- ITDMS thus is the largest data mining installation in India using non unique parameter like name and address.
- ITDMS is perhaps one of the largest data mining exercise in the world driven by any tax authorities using non unique parameter like name and address for search.

Awards and Recognition



Prime Minister Manmohan Singh (left) presents the Excellence Award in Public Administration to G.T. Venkateshwar Rao, IRS, additional commissioner, Income Tax, Hyderabad, for his role in development of Integrated Taxpayer Data Management system in New Delhi on Wednesday. — DC

PM awards IT officer for public service

DC CORRESPONDENT

NEW DELHI

April 21: The Prime Minister Dr Manmohan Singh on Wednesday gave away the awards for Excellence in Public Administration.

The additional commissioner, Income Tax, Hyderabad, Mr G.T. Venkateshwar Rao, was one of the recipients of the "Prime Minister's Awards for Excellence in Public Administration" for the year 2008-09. Mr Rao, along with his team members, received the award for preparing Integrated Taxpayer Data Management System programme.

TIMES NATION

THE TIMES OF INDIA, HYDERABAD | FRIDAY, FEBRUARY 6, 2009

HRD MINISTER SAYS THERE'S STILL SCOPE FOR MORE QUOTAS IN IIT, IIM POSTS | 9

SUPREME COURT EXTENDS DUTT'S ABROAD FOR A UN PROGRAMME |

Evaders Beware: I-T on the prowl

Taxmen Using '360-Degree' Profiling To Collect Details Of Tax Evaders

Pradeep Thakur | TNN

New Delhi: Officials of the air intelligence unit at Delhi's IGI Airport recently identified a traveller suspected to be carrying huge cash in his baggage. But before the suspect could be detained, he had boarded the aircraft for Hyderabad. All that the airline staff could divulge about the traveller were his initials and mobile number.

The number was soon flashed to the headquarters of the Central Information Bureau (CIB) of the Income-Tax department in Delhi. What happened thereafter is a startling example of how 360-degree profiling, a new cutting-edge database that synergizes with various agencies and service providers, is being used to nab income-tax evaders.

Within seconds of receiving the suspect's number, officials put it into their 360-degree data bank. It turned out to be that of a leading Delhi-based industrialist. In the next few minutes, sleuths had been dispatched to all business and residential premises of the accused. By the time the suspect landed at Hyderabad in a few hours' time, he was detained with the cash

A POTENT TOOL

What's 360-degree profiling?

It's a database of suspected money launderers, criminals and corrupt officials that tracks their activities and transactions in minute detail through synergy with various agencies and service providers

How does it map suspects?

Through their mobile conversations, credit card use, bank statements, travel tickets, hotel bills, real estate deals. Any transaction where PAN card is used shows up in suspect's profile



Who is under watch?

Some of the country's leading bureaucrats, politicians, journalists, lawyers and judges, industrialists and corporate houses

Has it helped recover money?

- Yes. A woman using three PANs, all registered at different addresses, was mapped and investigated last year. Within few days, her links with a leading real estate firm was established and she had to disclose hidden income of nearly Rs 100 crore
- Rs 40 crore recovered from Delhi-based stock broker on tip-off based on demat transactions

and confronted with all transactions made by him, his staff and family members in the past year.

Stunned by the extent of information officials had about his itinerary, contacts and transactions, the industrialist broke down and disclosed the source of the black money. He said it was meant to buy a property in Hyderabad.

Now increasingly being used by intelligence and enforcement agen-

cies, 360-degree profiles have turned out to be one of the most potent tools in creating forensic evidence against an offender providing minute-to-minute details of the movements of the accused. The details include all the people the target had spoken to, met, how much he spent with credit cards, where he was at a particular time, in which hotel did he check into and had how many visitors he entertained etc.

These profiles have been constructed on thousands of individuals and companies whose meteoric rise and flamboyance paint a suspicious picture around them. The I-T department already has financial details of these individuals and their front entities and family members. The profiles are constantly updated by sleuths based on analysis of annual information reports (AIRs) and other financial data received from several institutions, banks, registrar of societies, intelligence agencies, hotels and tour and travel operators.

The Central Information Bureau (CIB) of the Income Tax department is the nodal agency for collating this information. Different investigative agencies in the country take help of Central Bureau of Investigation and its data bank on 360-degree profiles while investigating a suspect.

The offices of the Central Bureau of Investigation offices are spread across the country situated in all major state capitals where local level intelligence and data received are analyzed and routinely fed into the system. Information related to banking, cash transactions and credit card expenses are centralized in Delhi.



India's taxmen get tech support to nail evaders

New software creates a complete taxpayer's profile and will dig out all information about a person on one screen

Thank You