SAMAGRA VEDIKA -- TELANGANA’S INTEGRATED PLATFORM

Using Big data, ML, Graph database

FOR BETTER CITIZEN SERVICE DELIVERY AND TRANSPARENCY, ACCOUNTABLE AND EFFICIENT GOVERNANCE

ITE&C DEPARTMENT
GOVERNMENT OF TELANGANA
Few of the welfare schemes of Telangana Total budget is more than 35,000 Cr

Ration Cards

Student scholarships

Aasara Pensions

Raithu Bandhu

Most of the welfare schemes have eligibility in terms of Income
<table>
<thead>
<tr>
<th>Types of Fraud</th>
<th>Leakage on account of</th>
<th>Resolution through</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Identity Fraud</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bogus Beneficiary</td>
<td>These leakages have been controlled through use of Aadhar &amp; ePOS</td>
</tr>
<tr>
<td></td>
<td>Non-existent real beneficiary</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Duplicate Beneficiary</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Multiple registrations by same beneficiary.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Right Beneficiary</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Limited means to establish identity of right beneficiaries.</td>
<td></td>
</tr>
<tr>
<td><strong>Quantity Fraud</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Illegitimate Claims</td>
<td>Research on mitigation is currently ongoing</td>
</tr>
<tr>
<td></td>
<td>Claiming bills in MNREGA without work</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Disproportionate Quantity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Availing more quantity of PDS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Food grains than eligibility</td>
<td></td>
</tr>
<tr>
<td><strong>Eligibility Fraud</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Is the person truly eligible?</td>
<td>No solution in the country as on date as it requires data of other departments</td>
</tr>
<tr>
<td></td>
<td>Limited means to correctly establish the eligibility</td>
<td></td>
</tr>
</tbody>
</table>
This is probably as big as a budget of a minor department!

Total Budget For 2019-20 For Pensions

₹10,000 Cr

Value of just 1% Leakage...

₹100 Crores
Eligibility for many benefits is prescribed.

People fulfilling one or more of the following conditions listed below shall not be eligible for Aasara Pension:

<table>
<thead>
<tr>
<th>Classification</th>
<th>Aspect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-Economic Indicators</td>
<td>Having land more than 3.0 acres wet/ irrigated dry or 7.5 acres dry.</td>
</tr>
<tr>
<td></td>
<td>Having large business Enterprise (oil/rice mills, pumps, shop owners etc.).</td>
</tr>
<tr>
<td></td>
<td>Owners of light and/or heavy automobiles (four wheelers and big vehicles)</td>
</tr>
<tr>
<td>Family Based Parameters</td>
<td>Having children who are Government/Public sector/ Private sector employment / Out-sourced/Contract.</td>
</tr>
<tr>
<td></td>
<td>Having children who are Doctors, Contractors, Professionals and Self employed.</td>
</tr>
<tr>
<td>Government Pensioner</td>
<td>Already receiving Government pensions or freedom fighter pensions.</td>
</tr>
<tr>
<td>Others</td>
<td>Any other criterion in which the verification officer may assess by the manner of lifestyle, occupation and possession of assets rendering the household as ineligible</td>
</tr>
</tbody>
</table>
Traditional process → ineffective implementation

Manual Process
- Reliance on Aadhar alone
- No Digital Trail

Effects of ignoring these challenges
- Inclusion Error (Benefit given to ineligible persons)
- Almost no accountability of officials in either error
- Exclusion errors (Denying eligible persons)

High Human Discretion
Opportunity & Challenges to get Consolidated view

1. Most of the Data is in electronic form
2. Data is in Silos
3. No Common ID
4. Integrated view – (SSOT) Single Source of Truth is not available
Samagra Vedika One View - Objectives

- **On-Demand Information**: On-demand information for informed decision making by authorized government officials.
- **Revenue Increase**: Property Tax, Water Tax, in Utilities.
- **Better Target of Subsidies**: Pension, PDS, Scholarships, subsidized housing, Health care.
- **Document Free Governance**: Document Free Service Delivery.
- **Law Enforcement**: Smart Policing.
What is the alternative approach
Without using Aadhar or any other ID
But getting the same efficacy In view of
Legal restrictions on use of Aadhar
Meta data attributes of an entity present in all data sources

- Following meta data information is available in every data source

  - Unique Personal Details
    - Name
    - DOB
    - Fathers Name (In Some)

  - Contact Details
    - Mobile No. (In some)
    - Address (Res)
    - Address (Off)

  - Unique ID* Number
    - PAN or
    - Passport no or
    - Voter ID or
    - Driving License

  *Any one ID is present

- All records in all data sources have Name, Address.
- Some records also have DoB, Phone Number, Fathers Name, Photo
- Can a combination of these attributes which are already available in every record be used to identify an entity
- With an Accuracy nearer to Aadhar based linkage
- With no manual intervention
3 “V” CHALLENGES

**VELOCITY**
The rate at which new data is generated is very high.

**VOLUME**
Large volume of existing data. About 38 Cr records.

**VARIETY**
No defined standards for writing name/address.

3V (Variety, Volume, Velocity) challenge in non-unique ID based record linkage is a classic Big Data use case.
### Examples of variations in Name & Father’s Name

<table>
<thead>
<tr>
<th></th>
<th>Variations</th>
</tr>
</thead>
</table>
| **Spelling**   | • N Radha Murali Krishna  
• N Radha Muralee Krishna |
| **Abbreviations** | • N R Murali Krishna  
• N R M Krishna |
| **Sequence Variation** | • Murali Krishna N Radha  
• N M Radha Krishna |
| **Addition/ Deletion** | • N Murali Krishna  
• Murali Krishna |
| **Splitting**  | • Murali Krishna N Radha  
• Radhakrishna N M |
Examples of variations in addresses

- Plot No 110 Rd No 65 J Hills Co Op Society Hyderabad
- P No 110 Rd No 65 Jubilee Hills Hyderabad
- Rd 65 Jubilee Hills Hyderabad
- 110 Road No 65 Jubilee Hills Hyderabad

Variations and errors in date inputs:
1. Interchange day and month: dd/mm/yyyy or mm/dd/yyyy
2. Input error in numbers
Key Performance Metrics

1. HIGH PRECISION
   High positive matching. Never miss the right match.

2. MINIMUM FALSE POSITIVES
   High negative matching accuracy. Get least number of false positives.

3. QUICK
   Very quick response time. Less than 2-3 seconds on 1 Cr records.

4. ABILITY TO RANK OUTPUT
   Ability to rank results so that the best results come on top.

5. DYNAMIC RULES
   Ability to define combination of multiple rules on Name, Address, DoB, etc.

6. HANDLE VARIATIONS
   Handle all variations in Name, Address, DoB etc.

All these are simultaneously achieved on very large databases (100 Million + records)
The following technologies are used:

1. Big Data Analysis Techniques
2. Entity Resolution
3. Graph DB: Relationship Discovery and Network Analysis
4. Machine Learning
5. AI Powered Image Recognition

The record linkage and entity resolution should be on non-unique attributes like Name, father's name, address, which vary greatly.
Economic Survey 2019 of GOI has praised Samagra Vedika

The Economic Survey 2018-19 singles out Telangana’s Samagra Vedika initiative as a shining example of digital transformation in governance. The policy document highlights the state’s efforts in digitizing government services, leveraging technology to improve efficiency and transparency. The initiative has won international recognition and has been praised for its innovative approach to data sharing and policy formulation.

Potential benefits
The Telangana Government’s Samagra Vedika initiative has been praised for its potential benefits. The initiative aims to integrate various government databases to provide a comprehensive view of the state’s economy and demographics. This integration will enable policymakers to make informed decisions and address development challenges more effectively.

Worst performers
But one parameter where the Telangana government didn’t shine was slum development. The state was ranked at 13th place by the survey, behind other states like Goa and Odisha.

While most states have reached 100% coverage of Indian Households Living in Poverty (IHLLP) at least on paper, Telangana lags way behind. The state has a slum population of 13.43%, which is much higher than the national average. This has serious implications for the state’s economic growth and social development.
Accurate targeting of subsidies

Beneficiaries for Old age pensions
Using Samagra Vedika for new sanctions of Aasara Pensions

Eligibility for Aasara Pensions is now 57 years

New applications are being received (expected about 7 to 8 lakh new pensioners)

65,693 new applications approved by the Districts officials after verification
Are sent to SERP for sanction Aasara Pensions

In Aug 2019 SERP requested ITEC to check the eligibility
Through Samagra Vedika platform

<table>
<thead>
<tr>
<th>Total new applications received</th>
<th>Total eligible applications (as per Samagra Vedika)</th>
<th>Total ineligible applications (as per Samagra Vedika)</th>
</tr>
</thead>
<tbody>
<tr>
<td>65,693</td>
<td>59,068</td>
<td>6,625 (10.1%)</td>
</tr>
</tbody>
</table>

Value Rs 16 Cr per annum
Accurate targeting of subsidies

Predictive analytics based identification of beneficiaries for 2 BHK scheme using Big data
Govt of Telangana has a program to provide 2BHK houses to economically weaker sections

- Started in 2015 by the Govt. Of Telangana to provide 100% subsidized housing to the poor.
- No beneficiary contribution needed – one of its kind in India.
- Construction cost = 7-8 lakhs/house and total cost including land is 15-20 lakhs/house

Aligned with the objective of implementation of scheme in the entire state, Govt of Telangana is looking to distribute 2BHK houses to eligible persons.

In one district

- Total Houses constructed under the scheme: 2000
- Total applications received: 11,681

The significant expenditure by Govt, and high mismatch in number of applicants and available houses has necessitated a very careful approach towards allotting the houses to applicants.
Siddipet Dist. Administration had following key observations.

- It is difficult to correctly identify the right beneficiaries based on the information collected in the application form.
- It is difficult to correctly identify the right beneficiaries based on any other information available with the district administration.
- Manual system.

- Some of the applicants have received subsidized housing earlier, but are reapplying using a family members name or their name.

- Certain applicants have submitted low incomes certificates even though they are financially well off.

- Some applicants or their family members already own a house.
Predictive analytics using big data used following data sets available with Govt.

Information provided by applicant

- Name
- Fathers name
- Address
- Aadhar number
- Phone number
- Photo of the applicant
- Minor info.

Information about family members not provided in the application

Datasets available

- Electricity connection
- Water connection
- House and land database
- Old age pension schemes
- Vehicles database
- Ration card database

Common databases are matched with the provided info. and they are further analyzed to bring out valuable insights in the form of applicant categories.
The analysis using Samagra Vedika categorized the applicants in four categories as follows:

<table>
<thead>
<tr>
<th>Categories -&gt;</th>
<th>Category 1</th>
<th>Category 2</th>
<th>Category 3</th>
<th>Category 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification</td>
<td>Qualify</td>
<td>Qualify with verification</td>
<td>Consider as low priority</td>
<td>Don’t consider</td>
</tr>
<tr>
<td>Not financially well off</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>No housing benefits previously accepted</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>No other welfare schemes prior</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>From Siddipet - SKS</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Count (% of Total)</td>
<td>2363 (20.2%)</td>
<td>2678 (22.9%)</td>
<td>2181 (18.7%)</td>
<td>4459 (38.2%)</td>
</tr>
</tbody>
</table>
Pilot at Hyderabad in Aug 2016

In Hyderabad about 1,00,000 cards were removed in Aug 2016

There was some public resistance due to which people were asked to apply again. About 19,000 applied as on Dec 16, 14,000 cards were activated again after verification that the property is very small or the four wheeler is taken out of loan etc.

Net about 86,000 cards are removed from August 16.

Total subsidy saved is Rs 4.6 Cr every month from Aug 16 onwards.

The mistake of tagging the vehicle/house to a wrong person is less than 5% which shows the efficiency of the application
Office of the  
Chief Rationing Officer,  
Hyderabad  

Dated: 31-12-2016

To,  
Sri G.T Venkateshwar Rao,  
Commissioner, ESD,  
Hyderabad

Sub:- CS - CRO's Unit - Information called for - regarding deletion of cards - report submission - Reg.
Ref:- Email Received from Sri G.T Venkateshwar Rao, Commissioner, ESD, Hyderabad dt 29-12-2016.

****

I invite kind attention to the reference cited wherein it was requested to furnish the information in the proforma prescribed. The information is prepared as given below.

<table>
<thead>
<tr>
<th>Month</th>
<th>Ration cards deleted</th>
<th>Units in deleted cards</th>
<th>Ration cards restored during the month</th>
<th>Net Ration cards deleted</th>
<th>Units in deleted cards</th>
<th>Saving of subsidy in Rs for Rice/Wheat/Sugar for that month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aug-16</td>
<td>10346</td>
<td>22783</td>
<td>0</td>
<td>10346</td>
<td>22783</td>
<td>₹36,90,848</td>
</tr>
<tr>
<td>Sep-16</td>
<td>102402</td>
<td>350809</td>
<td>2710</td>
<td>99992</td>
<td>341326</td>
<td>₹5,53,27,212</td>
</tr>
<tr>
<td>Oct-16</td>
<td>39906</td>
<td>119843</td>
<td>5248</td>
<td>34600</td>
<td>100009</td>
<td>₹6,87,944</td>
</tr>
<tr>
<td>Nov-16</td>
<td>10055</td>
<td>25489</td>
<td>14529</td>
<td>0</td>
<td>0</td>
<td>₹24,02,298</td>
</tr>
<tr>
<td>Dec-16</td>
<td>0</td>
<td>0</td>
<td>2723</td>
<td>0</td>
<td>0</td>
<td>₹43,407</td>
</tr>
</tbody>
</table>

Sd/- B. Bala Maya Devi, I.A.S.,  
Chief Rationing Officer,  
Hyderabad.

Note:- Applications for restoration of FS cards are being examined by concerned ACSO / Area Inspector and FS cards are restored if found eligible.

// Attested //

Assistant Civil Supply Officer  
Hyderabad
THANK YOU

Any Thoughts or Ideas

??