

Authors



Rajesh Aggrawal, IAS
Secretary, IT
Government of Maharashtra
sec.it@maharashtra.gov.in



Devroop Dhar
Principal Consultant, DIT
devroop.dhar@semt.maharashtra.gov.in



Murtuza Karachiwala
Senior Consultant, DIT
murtuza.karachiwala@semt.maharashtra.gov.in



Vithaldas Prabhu
Senior Consultant, DIT
vithaldas.prabhu@semt.maharashtra.gov.in

Directorate of Information Technology
7th floor, Mantralaya Annex
Mumbai-32

1. Introduction

In the last twenty years, India has undergone great transformation of economic and regulatory structures. Policy reforms in this period have led to healthy regulation. However, despite these efforts, access to finance has remained scarce in rural India. By issuing Aadhaar Card to the residents, social and financial inclusion of the needy has become reality. Maharashtra has taken big steps in adopting eGovernance at various levels. After completing the phase I of the UID enrolments, it was time to think on the next steps to be taken very quickly wherever enrolments have been completed to a great extent, considering the local population. Since in Maharashtra, especially in some districts like Wardha, penetration of enrolment is very high, Dept. of Information technology, Govt. of Maharashtra –started implementing various initiatives around Aadhaar number such as State Resident Data Hub (SRDH), Android tablet based enrolment centre survey, SMS based Aadhaar Self seeding, Web based Aadhaar self seeding, Usable SRDH (U-SRDH) etc.

Some might wonder, what is seeding and how seeding is done, who is responsible for seeding Aadhaar with various dept. databases. Seeding is the process to incorporate the Aadhaar number in various departments' database. Primarily seeding Aadhaar is the responsibility of each dept. Seeding can be done by adopting various methods. Seeding process can be a tedious manual process, it can be partially automated or it can be completely automated using various algorithms in a software application.

2. Seeding Methodology

UID seeding is a multi-step process. We need to first match the names and other demographic data in both the databases (UID database as well as Beneficiary Database). Once we identify correct UID data we need to link both the databases with UID number and we seed the UID number in to the beneficiary database. This database can even be a simple Excel sheet or it can

be one of the RDBMS database table. Once this is done, we need to link the bank account with UID number and bank's poller needs to be updated with the bank account number and Aadhaar number. This step enables the Aadhaar Enabled Payments to beneficiaries. In Wardha district, not a single method was sufficient and hence multiple methods were adopted while seeding Aadhaar number with various schemes' databases. Wherever beneficiaries' names were maintained in hand written list, it was first digitised using simple excel sheet. Microsoft Excel 2003 has limitation of maximum 65536 rows, hence if the number of beneficiaries are more than that, then we must use higher version of MS Excel.

Manual Seeding: Manual seeding can be done using MS Excel or using SRDH's manual seeding feature. While doing seeding using above both the methods, one has to search the name of the beneficiary in the UID database and identify correct record which matches with beneficiary database. There may be the case that more than one record has same or similar name. In this case, we need to consider other demographic data like Full address, Pin code, relationship or date of birth.

Semi-automatic seeding: Maharashtra has fully functional SRDH which has manual and batch seeding function. While doing the batch seeding, one has to first prepare the data of beneficiaries in a template provided by SRDH system, wherein weight-age of each field needs to be provided. The weight-age can be decided based on how confident we are about that particular data field. That means, if we think that names of most of the beneficiaries are correctly spelled, then we can give higher weight-age and give lower weight-age to other data fields. This excel sheet needs to be uploaded in to SRDH. SRDH uses the fuzzy logic and brings forward the list of probable matches. Department concerned needs to identify correct match and select for

seeding with Aadhaar number. Template for the batch seeding is added below:

A	B	C	D	E	F	G	H	I	J
department	departmentResidentId	uid	name	dob	address	district	state	pincode	gender
				20	40	10	10		20
P-D-S	BPL-0001		Suresh Savai Ranga	6/1/1974	B-214/5, Govt Colony, pandra	Etatara	MAHARASHTRA	400051	

Screen shot of the Batch Seeding template

In the above screen shot it can be observed that total weight-age comes to 100 i.e. 100%. First row contains the column header, second row contains the weight-age and third row onwards department data is listed. Using the fuzzy logic, most probable matches were listed and from this list, user department can pick the correct match considering other demographic data and seed the UID with the beneficiary database.

There is also a need to identify residents as a family. This will help us in finding out how many members of the family have enrolled and received the Aadhaar number and is there a member who is eligible for any of the government scheme as beneficiary but not in the system and hence not getting the benefit. We can make use of either Ration Card database or Socio-Economic Survey Database. Socio-Economic survey data has got the added advantage over Ration Card database as Socio-Economic survey data has most up-to-date information about the family. In Wardha district, Maharashtra, Socio-Economic Survey was first digitized and that was used to identify members of the family. This process of matching and linking UID with beneficiary database helped greatly in identifying the beneficiaries who were left out of the system. If Ration card data is also digitized and up-to-date information is available, then that can also be used for identifying the members of the family.

Why identifying unit of family is important?

1. With the help of relationship details, matching miss-spelled names / differently named person can be identified and matching and linking can be done. Sometimes it is observed that Janaki is mentioned as Jankibai or Sham is written as Shyamrao or Shravani is written as sharwani. Here we can check the middle name and relationship, date of birth, address details and find out if the person is same even if it is spelled differently.

2. Identifying left out beneficiary and bringing him / her into the system and providing direct benefit becomes possible.



Aadhaar Matching and Linking in progress

3. Ghost beneficiaries can be identified.

It is also pertinent to mention here that, just matching names will not help, it is also important that we must identify the beneficiary personally after matching and linking the names digitally. Door to Door survey will help in completing this task of verification.

Use of external vendors and technology partners: It is observed that many government departments do not have dedicated technical staff, which is trained and confident in using the computers for digitizing the department's data. Department has to take help of experienced data entry operators or technology partners for automating the process or for developing related applications. It is also important to mention here that there should be a process of reporting the progress regularly so that monitoring happens properly and areas can be identified where push is required to speed up the progress.

Outsourcing the technology related activities to empanelled vendors will help the department in resolving the issue of non-availability of experienced technical staff.

Some of the residents who were bed ridden or could not move due to old age, could not visit the enrolment centre. District administration made special arrangements to enrol such beneficiaries at their door step of the beneficiary so that they get their Aadhaar number and becomes part of the system and direct benefit transfer becomes possible.

will also be large. In this case, using one excel sheet for a district may not be possible as it requires higher end computers. One easy solution would be to split the work Taluka level and then within Taluka, Tehsildar, Talathi and circle officer takes care of set of villages in their jurisdiction.

Using Bulk SMS – District administration has sent bulk SMS to all the beneficiaries where mobile numbers are available – instructing them to contact their bank branch and provide their UID number to seed with bank account. It is also important to call beneficiaries for physical verification of UID seeding using Micro ATMs or Finger print scanner to confirm seeding of UID in the beneficiary database.

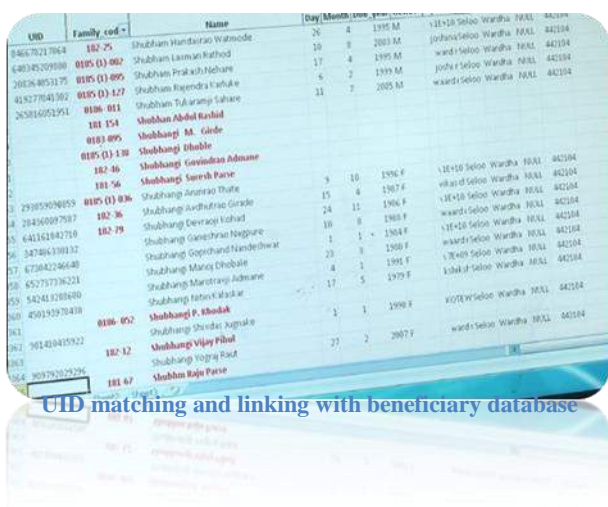
3. State Resident Data Hub (SRDH)

State Government of Maharashtra has prepared a State Resident Data Hub that houses the KYR information (name/gender/dob, etc) for every resident of Maharashtra who has registered for an Aadhaar (UID). Based on the registration data of the state/non-state residents, the SRDH application is a useful source of the KYR data and aims to leverage the data to the various State Departments Delivery Applications. Under a new initiative, Maharashtra State Government – Department of IT has prepared new services that will allow the resident user himself to link his KYR+ data to the Aadhaar details already available in SRDH. KYR+ information includes LPG Consumer No, PAN Card, Driving License, etc. Any State Resident user (Maharashtra) can seed his/her KYR+ information into SRDH database.

4. Self-Seeding – Features and usage

State Government of Maharashtra has prepared a State Resident Data Hub that houses the KYR information.

The various applications offered by the State empower the resident to seed his own data directly to the State Repository. No queues, no office staff, hassle-free way to directly interact with the State and update one's own data in the State Repository. State Resident User needs to send SMS to the Service Number (9222 - 200 - 022) in the pre-determined format. In case user does not send message in correct format, system will send a message with the correct format to the user.




How to handle large volume of data?


It is important to note that big districts might have large number of beneficiaries and also data for the UID

Prerequisite is that resident must have registered his mobile number in the UID database.


Resident can also scan the bar code to send the SMS, provided user has bar code scanner application is installed in the mobile.

Get your UID Number Mobile Number : 9222 - 200 - 022 SMS Text :MH SRDH MYUID	
---	---


[Get your UID Number by SMS](#)

Seed your Pan Card Number Mobile Number :9222 - 200 - 022 SMS Text :MH SRDH SEEDUID IT <Type PAN Card Number >	
---	---

[Seed PAN Card Number](#)

Seed your LPG Number Mobile Number :9222 - 200 - 022 SMS Text :MH SRDH SEEDUID LPG <Type Oil Company [IOCL/BPCL/HPCL]> <Type LPG Number >	
---	--

[Seed LPG Consumer Number](#)

Seed your Ration Card Number Mobile Number : 9222 - 200 - 022 SMS Text :MH SRDH SEEDUID PDS <Type Ration Card Number >	
---	---

[Seed Ration Card Number](#)

Self Seeding using Web Application



SRDH web based application can be used to seed resident data or Scan QR Code to Browse on Mobile. This option is alternate method to seed data if the user has access to internet

5. Issues with UID Data

During analysis of EID-UID data downloaded for SRDH, it was found that application used for enrolment had major issue of not having the masters in place for the villages and Talukas. Due to this, operators who have done enrolments had manually entered the data for village names and Taluka / Tehsil. One of the villages – “Sindhi Railway” was spelled in 14 different ways. This issue has been resolved in the next version of the application and is being used in second phase of UID enrolments. However existing data clean up is required.

It is observed that Borgaon named village exists in 6 different Tehsils and since Tehsil name is not captured while doing UID enrolment in the first phase, it is very difficult to match the resident data and link with beneficiary database. It is also observed that there were other issues like incorrect gender, photograph and some residents’ age was recorded more than 140 years, which is most probably incorrect. Specially constituted team to analyse the data found various such issues. Currently SRDH has data which needs cleaning by correcting the spelling of Village and Tehsil names, incorrect age, mismatch in gender, wrongly taken photograph etc. Understanding the difficulty, Dept of Information Technology, Government of Maharashtra visualised creating U-SRDH i.e. Usable SRDH.

6. Usable State Resident Data Hub (U-SRDH)

The implementation of the Direct Cash Transfer Scheme (DCTS) and other attempts for Aadhaar integration has encountered challenges due to lack of data standardization in the UID databases. In a unique initiative, the Directorate of Information Technology is currently undertaking a unparalleled data quality and data integration initiative to make the contents of SRDH usable. The initiative aims to create a Usable version of the SRDH (U-SRDH), which will have data (including address) completely standardized and cleaned.

Once standardized, the U-SRDH details would be leveraged for Aadhaar Integration across Central and State Government Schemes and eGovernance initiatives.

Recognizing that generic cleaning techniques and standardization rules would have limited impact given the specific localized context of Maharashtra State, through the U-SRDH the DIT is also undertaking several important initiatives for formulating rules through data mining and analysis.

A dedicated UID Laboratory has been set-up for the purpose of creation and maintenance of the U-SRDH. Some of the innovations and data processes currently underway at the UID laboratory include:

1. Data Mining for Name and address analysis for customized rule development

- The data mining process allows analysis of the raw data in the Temporary pre-staging area for development of customized rules for context specific data cleaning and standardization
- Here, scripts and algorithms are run on the raw data to provide analytical outputs and insights. The two types of analytical outputs expected are as follows:
 - Character and token frequency analysis for names and address fields of the UID details in the temporary pre-staging area
 - Character frequency analysis: The output gives frequency of occurrence of a particular character in at the start of, at the end of and in between the name field. Similar output should be generated for occurrence in the address field
 - Token frequency analysis: The output should be able to identify combinations of 2, 3 and more number of characters and report the frequency of occurrence of these combinations at the start of, at the end of and in between the name field. Similar output should be generated for occurrences in the address fields
 - Distance analysis for names and address fields: Once token analysis has been completed, the average distance between a character and / or a token would be computed for understanding typical occurrences and rules would be developed for treatment of outliers which have very high variance as compared to the average distance.

2. Standardization of name and address information in the U-SRDH through the customized rules developed by mining the data

3. Mining and Analysis to develop Marathi language competence and tools

4. Standardization and seeding of UID data with Election rolls and Ration card information

5. Standardization and seeding for beneficiary information from central and state schemes (especially for DCTS)

6. Standardization and seeding of household data from the Maharashtra State Electricity board and integration through GIS platform

7. Use of Know-Your-Resident (KYR) data for identification of false information declared under

several beneficiary schemes. For example, instances of potential wrong declaration of age for claiming scholarship was discovered, where age in KYR data was found to be 8 years, but the student was claiming scholarship for junior college

8. Experiments are currently underway to match UID photographs with the photographs in other datasets such as electoral rolls, driving license details etc.

7. Some examples of benefits through the U-SRDH

Seeding through the U-SRDH will enable validation of important information about beneficiaries under various Central and State Schemes. For example, research through the U-SRDH has shown that there is mismatch between the ages of student availing scholarships where U-SRDH KYR details have shown that the beneficiary is not old enough to avail that benefit. This could potentially be bogus beneficiaries.

Integration of data from various data sets, such as Ration card data, MSED data etc. would enable the department to get several insights which will enable individual departments validate each of these datasets. For example, assessment of village lists in election database should that there are several regions where a polling station may not be available. The UID laboratory is also planning to undertake assessment of correlation between voter turnouts and polling booth locations – especially in rural areas.

Matching of photographs could enable building further confidence into the in organic seeding done for Aadhaar integration across data sets.

The U-SRDH will enable rich analysis of data in the specific local context of Maharashtra State and will enable innovations for development of Marathi Language tools and competencies.

8. Innovation

Bed ridden and old age beneficiaries were not in a position to go out of their residence. Middlemen used to exploit the situation to their benefit and they were not at all getting the benefit of the Government schemes like old age pension, pension to handicapped persons, widow pension etc.

Technological innovation was used to overcome this problem. Each resident was issued UID number by capturing the biometric details like 10 finger prints, retina scan and photograph. Bank accounts were linked with the UID number and beneficiaries' grants were transferred to bank accounts linked with UID.

Business Correspondents (BC) were appointed by banks by providing Micro ATMs. BCs are required to visit allocated villages frequently and whoever needed to carry out banking transactions like cash deposit or withdrawal are being done right at the door step of the resident. While doing banking transaction, account holder needs to authenticate against the UID's central server where each resident's biometric data is stored and updated.

By UID authentication, beneficiaries started getting their full benefit right at the doorstep.

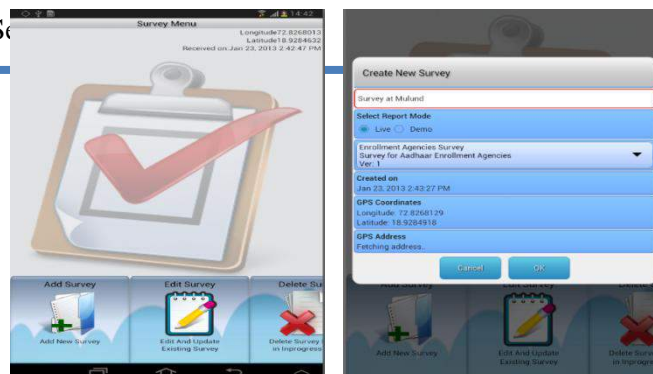
Various latest technologies were used e.g. Capturing biometric data like Finger prints and Retina scan.

Banking systems were linked with UID, UID Enabled payment bridge was created and hand held devices like Micro ATMs have been used as part of the solution. To complete the banking transactions successfully, robust connectivity has been established along with secured connection with the UID database.

For linking the UID with beneficiary data & bank account, in addition to innovations like SMS based self seeding of UID, Web based self seeding of UID with beneficiary database was carried out. Android based Tablet app was developed for doing the field survey of enrolment centres.

9. Android App for Enrolment centre survey

The State Government has implemented a tablet based application for conducting audit of UID enrolment centres and has made it available to other states over cloud. This type of monitoring of enrolment centre was very much required to have the quality check of enrolment data at the field level itself. Ensuring the quality of data at first instance itself is very much important.



This application has special feature –that is to configure the survey questionnaire as per the requirement of each survey. Currently this survey is being used by more than 5 departments such as UID enrolment, Survey for School Education, GSDA, Forest Dept, door to door UID seeding etc.

One more tablet app was rolled out for Door to Door seeding of UID with various other beneficiary databases.

10. Key Learning

While rolling out any technology initiatives covering large population, proper proof of concept needs to be rolled out first and data sampling is very much required. For any application, we need to have finalized master list such as State / district / Taluka/ village master, department master etc. In case of seeding UID with various beneficiary databases, only matching, linking and seeding on paper is not at all sufficient, we must also do the physical verification by any of the authentication mechanisms like biometric capture and verification or demographic data verification at the field. Banks also need to take care while seeding UID with the bank account. Department concerned is primarily responsible for seeding of UID with beneficiary database. The process needs to be streamlined for regular updates to the beneficiary database and upload for the direct benefit transfer.