Optimizing Agency Efficiency for Aadhaar Enrolments using Data Analytics





DATAWISE
Intelligence. Insights. Infinity.

Datawise
Hyderabad, India

info@mydatawise.com

+91 40 40204837

Aadhaar project is one of the important Mission Mode Projects (MMPs) that is considered as most important in creating overall efficiencies in the identification of individuals and ensuring that the services and benefits reach the intended audience with least leakages.

The enrolment process has now completed two years, and there is potential opportunity for improving the process of selection and enrolment of Agencies. Aadhaar relies on agencies to enroll individuals whose data is captured and stored in the database after cleansing for possible errors. Rejection of enrolment data results in loss of revenue and effort for the Agent, and has the potentiality to make the process unviable. A recent study has suggested that the internal rate of return from Aadhaar is likely to be in excess of 52%; however this is contingent on efficient enrolment process.

This study examines current enrolment efficiencies for agencies and suggests alternatives for improving the measurement efficiency of agency performance in order to make the process more dynamic.

1

BACKGROUND

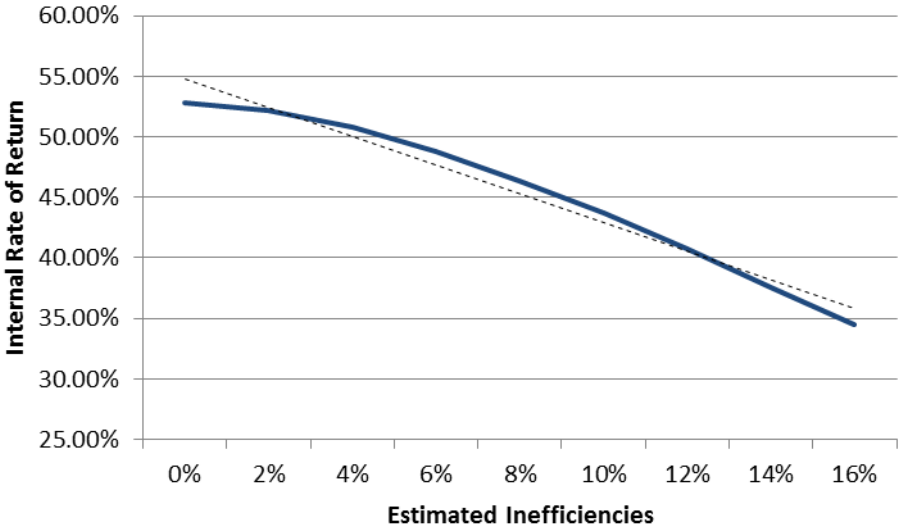
One of the most ambitious projects launched by the Government of India under the Mission Mode Projects (MMP) is the Aadhaar Project, also referred to as the UID (Unique Identification) project. The UID Authority of India (UIDAI) has been setup by the Govt. of India with a mandate to issue a unique identification number to all the residents in the country called as Aadhaar. UIDAI has a broad and ambitious mandate – issuing unique identity numbers to the 1.2 billion residents in India. The primary goal of Aadhaar is to identify all individuals based on a set of 12 parameters that have been predetermined, and which together are expected to uniquely identify each individual.

In recent times, the process of enrollment under Aadhaar has faced some controversies during the last few months, especially with concerns about cost and time overruns. On 27 January 2012 the Cabinet Committee on UIDAI announced that UIDAI will be allowed to enroll additional 40 crores million residents beyond 20 crores initially approved. While this has breathed life into the Aadhaar program once again, there is a likelihood of far greater scrutiny of UIDAI especially since the beneficial outcome from the Aadhaar enrolment program cannot be expected until a reasonably large number of persons are enrolled.

[1] This Paper is a summarized version of the Paper presented by Datawise at IIM Ahmedabad by K Vinay Kumar.

A recent paper on the Cost Benefit analysis (A cost-benefit analysis of Aadhaar, November 2012) of deploying Aadhaar has estimated that the potential rate of real returns in real terms to the Government of India is 52.85 percent. This is based on an estimated cost of enrolment of about Rs 16,386 crores spread over a period of 11 years from 2010-11 to 2020-21. This translates to an approximate cost of Rs 130 per enrolment. The real benefits notwithstanding, any inefficiencies in the process of enrolment due to increased rejections could push up the total cost of enrolment which could impact the estimated rate of return. The incremental impact of inefficiencies on the internal rate of return for Aadhaar is as shown in the Figure below:

Figure 1: Impact of inefficiencies on Internal Rate of Return (IRR) for Aadhaar Project



UIDAI has developed an algorithm that ensures authentication of data. Through this process, the UIDAI only enrolls residents after de-duplicating records. This shall help Registrars clean out duplicates from their databases, enabling significant efficiencies and cost savings. The strong authentication that the UID number offers is expected to improve services, leading to better resident satisfaction.

Eliminating duplication under various schemes is expected to save the government exchequer upwards of Rs. 20,000 crores a year. It will also provide governments with accurate data on residents, enable direct benefit programs, and allow government departments to coordinate investments and share information.

Enrolments under the Aadhaar program were successfully launched on 29 September 2010; the program has currently enrolled about 27.79 crore persons from 625 districts all across India. The enrolment process had been awarded to 150 agencies who directly work with registrars to complete the enrolment process. Of these, in the year 2012, only 112 have conducted Aadhaar enrolments, indicating a drop out of agencies (a drop-out rate of 34%).

Agencies were selected based on a prequalification criterion determined by UIDA, based on which they were further classified to an expected capacity performance. Agencies were classified based on Technical Capability (T1 to T4) and Financial Capability (F1 to F4). Agencies that had demonstrated interest would automatically be classified as T1 (entry level); and graduate to T2 on employing 25

enrollers, or completing 50,000 enrolments; graduate to T3 on employing 75 enrollers, and completing 10 Lakhs enrolments.

The financial capabilities too were judged on the basis of Net worth (Rs.50 lakhs to Rs 2 crores for F1, Rs. 2 crores to Rs. 5 crores for F2, Rs. 5 crores to Rs. 10 crores for F3, and above Rs. 20 crore for F4). The broad classification of agencies under each classification and its implications are as follows:

Table 1: Eligibility of agencies based on classification

Tier	No. of States eligible	Maximum number of enrolments in a year	Number of Agencies
F1	2 states	15 lakhs	30
F2	4 states	35 lakhs	26
F3	8 states	125 lakhs	36
F4	Any number	500 lakhs	54

In theory, if each agency does its maximum level of enrolments, the total enrolments possible in a year are 32.86 crores. However, in the current year so far, the total number of enrollments have only been about 11 crores, clearly indicating a significant gap between expected rate of enrolment and the actual enrolment.

It should be expected that the higher is the classification of an agency, the better would be its performance in terms of enrolments as well as efficiencies.

Of the 112 agencies that have conducted enrolments during the year 2012, at least 46 have not been assigned any classification. We believe that this is possible due to a lag in the updating data on public websites.

AGENCY PERFORMANCE

The success of the enrolment program is primarily dependent on the ability of the agencies to complete the enrolment process efficiently. The enrolment agency's performance is dependent on the quality of its infrastructure, staff, enrolment processes.

Once an enrolment agency submits the data relating to a person, Central Identities Data Repository (CIDR) carries out the necessary quality and validation checks, following which a de-duplication algorithm is run to ensure uniqueness. Any cases where process data errors, or duplicates are detected, they are rejected and not added to the database.

AGENCY EFFICIENCY

From an enrolment agency's perspective, the greater the number of enrolments, the higher is the remuneration that they receive. Most agencies have optimized their business model to the extent that the relationship between their revenues and costs is closely aligned. As a result, their downside risks

on poorer rates of enrolment are lower than the risks of not achieving the desired result for the Government.

However, every rejection is likely to increase agency risks due to:

1. Lower payments
2. Risk of being suspended

From UIDAI's perspective, agency efficiency is best measured based on multiple parameters which include rate of new enrolments, rate of rejections, ratio of rejections to enrolments, benchmarked performance to agencies within the district, benchmarked performance to other agencies within the same tier, and so on.

As an implementing body, it serves UIDAI's interests best if it is able to establish a robust mechanism of measuring, monitoring, and weeding out poor performing agencies. Setting up a dynamic system of measurement and monitoring (as opposed to a static, periodic review) will help UIDAI be more responsive and ensure efficient operations. This paper presents an analytical process and framework for measuring agency performance on a daily basis which can be used for effective agency monitoring and control.

DATA FOR ANALYTICS

UIDAI generates standard reports for daily monitoring of process performance which is collated and available on its public portal. This data is updated on a daily basis based on information collected across all agencies performance across all districts. This data is available across multiple parameters such as date, Registrar, Agency, State, City, District, and Pin Code. For each date, the number of enrolments, and rejections are available and have been used for analysis of agency efficiency.

Aadhaar data was extracted for the period from January 2012 to December 2012. The data was initially extracted in CSV format and input into infobright, which uses columnar approach to database design. While loading the data into a table, it is broken into the groups of 2^{16} rows, further decomposed into separate data packs for each of the columns. This makes the search and querying of big data easier, since columns are compared rather than rows. The input CSV files which were based on date had following columns.

In these following columns were considered as inputs for database

- Registrar
- Enrolment Agency
- State
- District
- Sub District
- Aadhaar generated
- Enrolment rejected

In addition to the above mentioned fields filename which had the date details was also considered. The data was fed infobright database, which in total summed up to 51,076,352 rows The field "Aadhaar Generated" gives enrolments on the given date for a given enrolment agency. The field

"enrollment rejected" gives the rejections on a given date for a given agency. The above fields were inserted into a table Aadhaar_info_2012.

The metadata of enrollment agency, namely enrollment agency code, name of agency, contact person, capacity per rfe1, capacity per rfe2, and state allotted were taken into a separate table Aadhaar_metadata. For obtaining the final results, these two tables were joined using a common column as reference. According to rules of relational database, we used 3rd degree of normalization to normalize the tables. Through this joining query, duplication of data was avoided taking into consideration the need to ensure data integrity.

A preliminary review of the data revealed at least 223 instances during the 12-month period from 1 January 2012 to 1 January 2013, where the total rejections for an enrolling agency were higher than the total enrolments during the period. We attribute most of these to the time lag in detecting errors during the enrolment process. However, there were some cases where the data appeared clearly incorrect (for example, for Clairvoyance Technologies, total rejections in Nagaland were 52,448 against 72 enrolments during the same period). All such cases were eliminated for the purpose of analysis.

"A preliminary review of the data revealed at least 223 instances during the 12-month period from 1 January 2012 to 1 January 2013, where the total rejections for an enrolling agency were higher than the total enrolments during the period. We attribute most of these to the time lag in detecting errors during the enrolment process."

PREDICTING AGENCY PERFORMANCE EFFICIENCY

The premise for creating Technical and Financial Capability based classification presupposes that the ability of an agency to perform its role depends to a large extent on this classification. We therefore attempted to determine whether there was any correlation between the classification of the agenciesand their performance. An analysis of these performances based on their Financial Classification shows that if there was any correlation at all, it was negative, that is, the higher the classification of an agency, the greater is its average rate of rejection.

Table 2: Relationship between Financial Capabilities and Agency Efficiency

Financial Capabilities	Average Rejection rate
F1	0.9%
F2	1.0%
F3	1.9%
F4	2.3%

There was no evident relation between Technical Capabilities classification and the rates of rejection of agencies. Evidently, the capabilities of an agency based on their size and capacity determine do not determine their efficiency. As a result, it became necessary to examine alternate reasons for poorer efficiencies for enrolling agencies.

ANALYSIS

Of the 630 districts, as many as 30 districts had a rejection rate (calculated as the number of total rejections in the 12 month period from 1 January 2012 to 31 December 2012 divided by the total enrolments in each of the districts) in excess of 10%. This is potentially alarming because of its implications for agency efficiencies. Further, 59 districts demonstrated a rejection rate of 5% or higher (9% of the total districts). A summary of this is shown in the Table below:

Table 3: Geographical dispersion of low efficiency agencies

State	Total number of agencies	Districts with more than 10% Rejection	Total Number of Districts	% of high rejection districts*
Andhra Pradesh	88	4	23	17%
Chhattisgarh	82	3	22	14%
Goa	45	1	3	33%
Kerala	79	3	14	21%
Lakshadweep	17	1	1	100%
Manipur	58	1	9	11%
Nagaland	43	5	11	45%
Rajasthan	84	1	33	3%
Tamil Nadu	77	4	32	13%
Uttar Pradesh	92	1	70	1%
Uttarakhand	79	1	13	8%
West Bengal	91	3	19	16%

** For districts with 10% or excess rejection rates*

Twelve States have districts which have high rejection rates – of these, three States (Lakshadweep, Manipur, Nagaland) are difficult states. A bigger cause for concern, however, is that 5 of the states that have high rejection rates (Tamil Nadu, Andhra Pradesh, Kerala, West Bengal, and Chhattisgarh) are all high population states, which together account for more than 300 million people. There was however no demonstrable correlation between the population of a state and the rate of rejection. Even if one were to examine the rate of enrolment compared to the population of the district, no indications as to the efficiency of the agency emerged.

However, a further analysis of the performance within these states and districts revealed that a majority of the rejections were from a few Agencies that accounted for a majority of the rejections. The details of these agencies and the proportion of the total rejections in the district that they represent are as follows:

Table 4: Agencies with high rejection rates in poor performance geographies

Sl. No	Agency	Total rejections*	Total enrolments*	% Rejection
1.	Swathy Smartcards Hi-Tech Pvt	520,979	9,766,365	5.3%
2.	Swisstech NPR 57cr Project Pvt	371,730	791,079	47.0%
3.	Pioneer E Labs limited	153,506	602,509	25.5%
4.	Madras Security Printers Ltd^	81,101	2,153,005	3.8%
5.	Eagle Software India Pvt. Ltd	70,335	419,036	16.8%

Sl. No	Agency	Total rejections*	Total enrolments*	% Rejection
6.	The Peerless General Finance	27,325	329,936	8.3%
7.	India Computer Technology	19,422	683,870	2.8%
8.	Datasoft Computer Services(P)	8,330	1,124,934	0.7%
9.	Vakrangee Softwares Limited	7,892	8,007,331	0.1%
10.	Smart Chip Limited	100	3,305,688	0.0%
11.	Computer LAB	41	4,063,086	0.0%

*Note: * During the calendar year 2012*
^According to a letter dated 23 August 2011, this agency's empanelment has not been renewed for the year 2011-12, and yet, data was available from the portal

It is noteworthy that of the total nearly 77 lakh rejections during the year 2012, the abovementioned 11 agencies accounted for 16.5% of the total rejections.

If one were to examine the rejections based on which agencies have been the greatest offenders irrespective of geography, the trend is not very dissimilar. The table below shows the details of enrolments and rejections for the agencies with the highest rejections (in excess of 100,000) during the year:

Table 5: Agencies with high overall rejection rates

Sl. No.	Agency	Enrolments*	Rejections*	Rejections %
1.	Swathy Smartcards Hi-Tech Pvt	9,766,216	1,708,738	17%
2.	Tera Software Ltd	5,396,544	669,971	12%
3.	Swisstech NPR 57cr Project Pvt	791,063	626,877	79%
4.	Computer LAB	4,062,865	377,007	9%
5.	Vakrangee Softwares Limited	8,002,646	366,618	5%
6.	Emdee Digitronics Pvt.Ltd.	1,418,623	292,906	21%
7.	Madras Security Printers Ltd^	2,151,797	288,037	13%
8.	Multiwave Innovation	1,036,777	203,191	20%
9.	Wipro Ltd	7,441,652	177,152	2%
10.	Krishna Infotech	250,512	176,385	70%
11.	Pioneer E Labs limited	600,855	154,733	26%
12.	Akshaya	6,812,346	153,153	2%
13.	Virgo Softech Limited	5,508,291	151,370	3%
14.	TechSmart India Pvt Ltd	2,192,967	137,787	6%
15.	In Media Computer Services LLP	1,272,958	129,553	10%
16.	Eagle Software India Pvt. Ltd	416,463	112,280	27%

*Note: * During the calendar year 2012*
^According to a letter dated 23 August 2011, this agency's empanelment has not been renewed for the year 2011-12, and yet, data was available from the portal

These 16 agencies together accounted for 57 lakh out of the 76 lakh rejections during 2012, representing 75% of the overall rejections. Given the high concentration of rejection either within agencies or within specific geographies, there is a case for closer examination of the causes, and possible identification of reasons for such performance. Six agencies are common to both the above lists, and require greater scrutiny. These are as shown in the table below:

Table 6: Worst performing Agencies

Agency Name	Financial Classification	Technical Classification
Swathy Smartcards Hi-Tech Pvt	F3	T1
Swisstech NPR 57cr Project Pvt	N.A.	N.A
Pioneer E Labs Limited	N.A.	N.A
Eagle Software India Pvt. Ltd	F3	T3
Vakrangee Softwares Limited	F4	T3
Computer LAB	N.A.	N.A

CONCLUSIONS

Our analysis shows that the rate of rejections for some of the agencies that have had lower enrolments is high. In fact, based on analysis of the data for the year 2012, there exists a weak negative correlation between the number of enrolments and the rejection percentages.

Based on our analysis, there are at least 12 agencies which have a rejection rate in excess of 10%, clearly indicating inefficient operations. These include Swisstech NPR 57cr Project Pvt, Krishna Infotech, Eagle Software India Pvt. Ltd, Pioneer E Labs limited, Emdee Digitronics Pvt.Ltd., Multiwave Innovation, Swathy Smartcards Hi-Tech Pvt, Vedavaag Systems Limited, Madras Security Printers Ltd, Tera Software Ltd, In Media Computer Services LLP, and India Computer Technology. These require to be replaced by more efficient agencies that can improve net enrolments.

We also discovered a lag between the dates of enrolment and rejection. On an average, this was between 5 to 8 months. As a result of this, inefficient operations of an agent remain undetected for a long period of time. This delay in measurement of agency efficiency will have an overall impact on the program's ability to meet its targeted enrolments in a timely manner.

It is therefore critical for UIDAI to reexamine the following:

1. The criteria for empanelment of agencies has to be thoroughly re-examined. While the technical and financial capabilities of the agency (which are important for selection) appear to have no bearing on the performance of the agency, we suspect that a greater role is played by the pricing that is proposed by the agency at the time of submitting its bids to the registrar for enrolment.
2. The current process of monitoring, which relies on periodic audits, results in significant delays between inefficient enrolment procedures, and their detection. As a result, the overall cost of the enrolment process and the viability of the Aadhaar project is likely to be affected. Since enrolment and rejection data is readily available, this should be used for setting up a continuous monitoring process to weed out the poor performers and ensure greater success of the project.
3. For future empanelment, an alternative method of agency selection should be selected, which should, among others, include a measure of the efficiency of the operator, and not rely solely on the number of enrolments for technical capability classification.

About DATAWISE®

DATAWISE® offers a suite of products and solutions suited to the needs of various situations and industries. Solutions provided for one customer are not necessarily suitable for others, and readers are advised to use their own judgment regarding the suitability of these solutions to their business needs.

DATAWISE®'s business analysis services support the full spectrum of clients' needs with services directed mainly at helping companies discover opportunities for improvement through use of analytical capabilities. We offer analytical services in the following areas:

Strategic Analytics: Alignment of strategic intent with actual work, requiring strategic analytics to answer key decision support questions such as whether to enter into a new segment of business or not, whether to reach new customers or not, and other go, no-go decisions.

Behavioral Analytics: Assistance in determining the 'why' and 'how' of a customer behavior (rather than the 'what') in order to ensure that marketing plans yield the desired results through capturing customer events and actions over time and using these stored interactions to determine typical behavior and deviations from that behavior.

Tactical Analytics: Tactical analytics models that we deploy are typically short-term in nature, and are focused on answering immediate questions rather than aligning to a longer-term goal.

Predictive Analytics: We created complex multi-dimensional models that collate data generated from several interaction points to create models that enable the prediction of future events to help identify of both risks and opportunities.

DATAWISE® has also developed proprietary analytics models OPTLIOX™, CREST™, Infinity™ and DATTAB™, catering to specific customer needs.

• Hyderabad • Delhi • Mumbai • Bangalore • Jaipur • USA

www.mydatawise.com mail at info@mydatawise.com