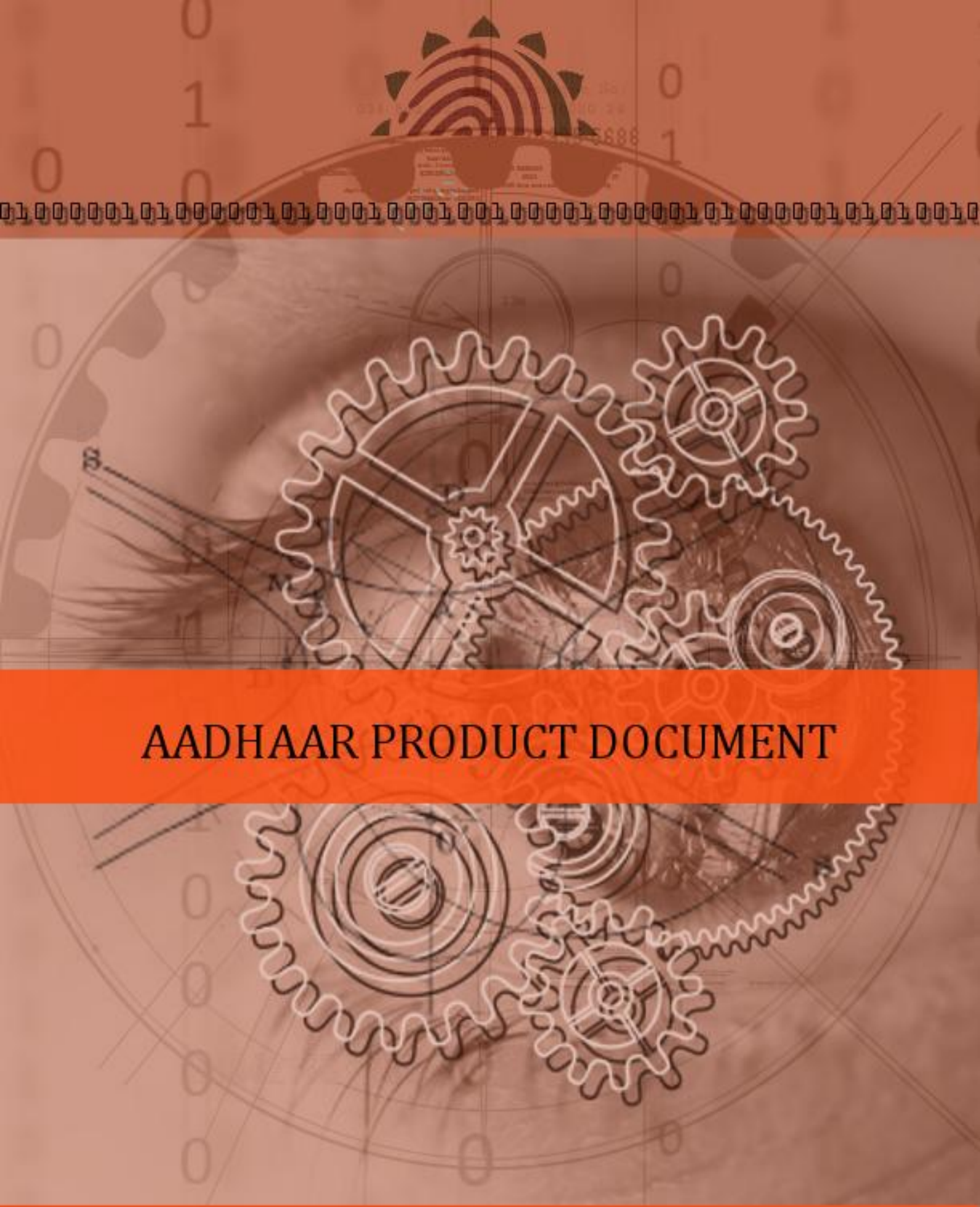


AADHAAR PRODUCT DOCUMENT

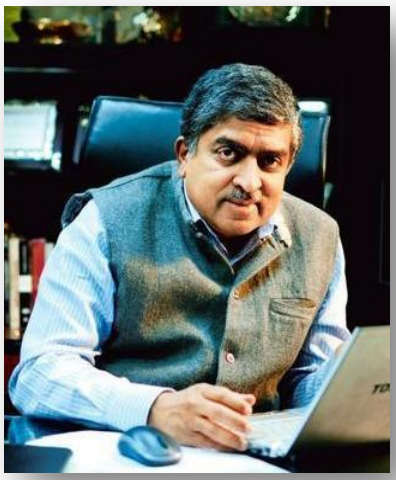
UIDAI Technology Center
March 2014

AADHAAR PRODUCT DOCUMENT

MARCH 2014

UNIQUE IDENTIFICATION AUTHORITY OF INDIA
PLANNING COMMISSION, GOVERNMENT OF INDIA
JEEVAN BHARATI BUILDING, CONNAUGHT CIRCUS, NEW DELHI – 110001

Message from Chairman



UIDAI has the vision of empowering every resident of India with a unique identity and providing a digital platform to authenticate anytime anywhere. The Aadhaar system is built on a sound strategy and a strong technology backbone and has evolved into a vital digital infrastructure. Meticulous planning and execution enabled the program to be launched ahead of plan in Sept 2010 and reach the kind of scale that was never achieved in any biometric identity system in the world.

UIDAI created an ecosystem of partners for the various key components of the project and integrated them all onto a common technology backbone using open standards. Aadhaar technology system was able to succeed due to its core principles - openness, vendor-neutrality, security, and data analytics. In the absence of digital privacy laws, UIDAI took it upon itself to implement rigorous standards and measures to ensure data privacy and security. Apart from generating 60 crore (600million) Aadhaars in 4 years through this approach, the Aadhaar platform is now integrated into the financial systems of NPCI and banks, taking India towards the goal of total financial inclusion.

Documenting of the design and implementation of the Aadhaar technology, which evolved over a period of four years, was as onerous as building the system itself. The UIDAI Technology Centre has done a commendable job in compiling white papers on technology strategy, application features, and architecture. These documents place the Aadhaar system on a firm foundation and also serve as the beacons for many e-Governance projects in India and across the world.

My hearty congratulations to the UIDAI technology team!

Nandan Nilekani
Chairman, UIDAI

Message from Director General

Aadhaar project, under the Planning Commission, Government of India, is an initiative to provide a unique identification number to every resident that can be leveraged by the residents to access various services and benefits. Uniqueness of Aadhaar identity allows elimination of fake and duplicate accounts, and online authentication provides a mechanism for paperless, electronic and instantaneous verification of identity, anytime anywhere. Aadhaar platform can be utilized by various Governmental, public, and private sector agencies to efficiently deliver services to residents.

By its very nature, the Aadhaar system needed a strong technology foundation. Appropriately, the Technology Centre was the first unit of UIDAI to start functioning in 2009 in Bangalore. UIDAI technology team was able to develop the required applications and successfully generate the first Aadhaar number within a year of its formation. The architecture was rigorously tested and validated through a series of 'Proof of Concept' studies. Aadhaar generation was commenced in August 2010 and the first Aadhaar letter was formally handed over to a resident at Nandurbar in Maharashtra on 29 September 2010. Subsequently Authentication and e-KYC services were also launched.

In the course of attaining the milestone of 60 crore (600 million), the Aadhaar technology backend has become the largest biometric identity repository in the world and the first to provide an online, anytime anywhere, multi-factor authentication service. A strong technology foundation based on open architecture enabled the rapid evolution of the Aadhaar system. It was important to document all aspects of Aadhaar technology and make it available in public domain. The three white papers published by the UIDAI Technology Centre fulfil this need.

I sincerely appreciate the efforts of our technology team in publishing these.

Vijay Madan
Director General, UIDAI

From the Editor's Desk

'Aadhaar' is undoubtedly one of the most important Projects rolled out by Government of India. Ambitious, one-of-a-kind and a game-changer, 'Aadhaar' is on the path of delivery to every Indian resident, a 'national identity' and triggering thereby the much desired governance system based on social inclusion, transparency & accountability.

Considering India's population of 121 crores, it is obviously a highly ambitious project. Any Project of this magnitude prospers and grows based on the contributions from the people associated with it. Unique Identification Authority of India (UIDAI) has been blessed in this regard. Many outstanding people from both Private and Government sectors, spanning the domains of - Administration, Biometrics, Project Management, Law, Technology and Finance, to name a few - have come together and dedicated their talent, time and knowledge, besides the much needed passion & commitment.

UIDAI has also been led by competent people like Shri Nandan Nilekani, as its Chairman, Shri Ram Sewak Sharma, as its first Director General and Dr Vijay Madan, who succeeded him a year back. Shri Nilekani has provided both visionary and charismatic leadership to the Organisation as its Founding Chairman. With his experience of building one of India's internationally recognised I.T. Companies; besides his association with both State & Union Governments in different advisory capacities, he was an appropriate choice to establish and lead an organisation like UIDAI that called for an understanding of technology and a commitment to social inclusion. The yeoman services rendered by the first Director General, Shri R.S Sharma are recalled with fondness. A man of boundless energy and dedication, Shri Sharma with his attention to detail and sharp technical knowledge, ensured that the Project got the right kick-start and maintained momentum in its formative years. Dr Madan has been the perfect foil to the Chairman besides being an able successor as the current Director General. He has been able to expand the horizon of the Organisation by steering it to the next growth trajectory for quicker delivery of Aadhaar and its deployment for people-centric Applications. His focus also has

been on knitting together different components into a single and synergetic unit for this purpose.

UIDAI has always believed in documentation and sharing of information with public. Technology & processes that characterise UIDAI needed to be brought within the public domain as a matter of transparency, as also to elicit debate and discourse. Towards this end, the White Papers have been in the making for a while. I acknowledge the contributions of Shri Srikanth Nadhamuni, Dr Pramod Varma, Shri Sanjay Jain and Dr Vivek Raghavan in writing the papers. Our ADGs at the Tech Centre too have been actively involved in writing and reviewing them. The contributions of Shri Rajendra Kumar, Shri Sudhir Narayana, Shri Venkat Rao and Shri Anup Kumar are appreciated in particular.

The writers and reviewers have not lost sight of the need for a simple & straight language while putting together the highly technical and procedurally rigorous content of the Project. They have done a commendable job. Further, these documents are a manifestation of an intensely cerebral and immensely taxing effort put in by a band of dedicated domain experts who worked with the UIDAI's Tech Centre at different stages of its evolution. There are many, but some who definitely need a mention are Srikanth Nadhamuni, Raj Mashruwala, Pramod Varma, Sanjay Jain, Vivek Raghavan, and Jagdish Babu. Our colleagues including Dy. Directors General, Asst. Directors General and others serving UIDAI's cause at the Headquarters, Regional Offices and Technology Centre have contributed in multiple ways to the evolution of technology, processes & procedures and the compilation of this document. I beg pardon for not mentioning them by name. All of them are gratefully acknowledged.

It is a matter of pride and pleasure that these are getting published as the UIDAI reaches the hallmark of 60 crore Aadhaars well ahead of the targeted date and gears up to provide identification to the other half of the nation.

Ashok Dalwai
Deputy Director General, UIDAI
Technology Centre, Bengaluru

Table of Contents

MESSAGE FROM CHAIRMAN **5**

MESSAGE FROM DIRECTOR GENERAL **7**

FROM THE EDITOR'S DESK **9**

TABLE OF CONTENTS **11**

TABLE OF ABBREVIATIONS **15**

TABLE OF FIGURES **17**

PART I: INTRODUCTION & OVERVIEW **19**

1 INTRODUCTION **21**

1.1 PRIMARY FUNCTIONS 21
1.2 PRODUCT GOALS & PRINCIPLES 22
1.3 DESIGN CONSIDERATIONS 23
1.3.1 DESIGN FOR INCLUSION & CHOICE 23
1.3.2 ECOSYSTEM APPROACH 24
1.3.3 IDENTITY AS A PLATFORM 26
1.3.4 STANDARDIZED TECHNOLOGY AND PROCESSES 26
1.3.5 PRIVACY BY DESIGN 26
1.3.6 FEDERATED DATA MODEL & ONE WAY LINKAGE 29

2 THE AADHAAR SYSTEM **31**

2.1 THE AADHAAR APPROACH TO IDENTITY 31
2.2 FEATURES OF THE UIDAI MODEL 31
2.2.1 DEMOGRAPHIC DATA COLLECTION 32
2.2.2 BIOMETRIC DATA COLLECTION 32
2.3 THE AADHAAR SYSTEM 33
2.4 COMMON APPLICATION REQUIREMENTS 34
2.5 ENROLMENT 35
2.6 AUTHENTICATION 36
2.7 ELECTRONIC KYC 36
2.8 PLATFORM AND COMMON MODULES 37
2.9 FRAUD MANAGEMENT 37

3 INFORMATION PRIVACY & SECURITY **39**

3.1 INTRODUCTION 39
3.2 DATA WITHIN THE AADHAAR SYSTEM 39
3.3 PERSONAL IDENTIFIABLE INFORMATION (PII) 40
3.4 DATA RETENTION AND USAGE 41
3.5 SYSTEM SECURITY 42

4 SYSTEM INTERFACES **45**

4.1 ENROLMENT CLIENT 45
4.1.1 AADHAAR BIOMETRIC CAPTURE DEVICE 46
4.1.2 BIOMETRIC SDK 46
4.1.3 GPS 46

4.1.4 DOCUMENT SCANNER 47
4.1.5 TRANSLITERATION 47
4.1.6 MASTER DATA IMPORT 47
4.1.7 SECURE PACKET TRANSPORT, UPLOAD & MANAGEMENT 48
4.2 ENROLMENT SERVER 48
4.2.1 ABIS 48
4.2.2 LOGISTICS PARTNER INTERFACE 49
4.3 AUTHENTICATION API 49
4.4 E-KYC API 49

5 THE AADHAAR NUMBER 51

5.1 REQUIREMENTS 51
5.2 SOLUTION 51
5.3 FORMAT FOR DISPLAY AND PRINTING 52
5.4 RESERVED NUMBERS 52
5.5 RANDOM NUMBER GENERATOR 52

PART II: ENROLMENT 53

6 RESIDENT ENROLMENT 55

6.1 THE AADHAAR ENROLMENT PROCESS 55
6.1.1 RESIDENT IDENTITY VERIFICATION 56
6.1.2 DEMOGRAPHIC QUALITY CHECKS 56
6.1.3 OPERATOR / SUPERVISOR VERIFICATION 56
6.1.4 EXCEPTION MANAGEMENT 57
6.1.5 FIELD MONITORING AND QUALITY MANAGEMENT 57
6.2 THE ENROLMENT ECOSYSTEM 58
6.2.1 UIDAI 59
6.2.2 REGISTRARS 59
6.2.3 ENROLMENT AGENCIES 59
6.2.4 OPERATORS / SUPERVISORS 59
6.2.5 TRAINING DEVELOPMENT AGENCIES 60
6.2.6 OPERATOR TESTING & CERTIFICATION AGENCIES 60
6.2.7 BIOMETRIC DEVICE CERTIFICATION 60
6.3 ENROLMENT SYSTEM REQUIREMENTS 60
6.3.1 LANGUAGE SUPPORT 60
6.3.2 EXTERNAL ARTEFACTS & DOCUMENTS FOR RESIDENTS 61
6.4 DEVICE ECOSYSTEM 62
6.5 ENROLMENT MODULE 63
6.6 ENROLMENT CLIENT 63
6.6.1 CLIENT FEATURES 64
6.6.2 DEVICE SUPPORT 64
6.6.3 CLIENT SECURITY 64
6.6.4 BIOMETRICS SUBSYSTEM 66
6.6.5 CLIENT MASTER DATA 69
6.6.6 DEMOGRAPHIC DATA VALIDATION 70
6.6.7 ENROLMENT ID 71
6.6.8 RESIDENT DATA CORRECTION 73
6.6.9 ENROLMENT PACKET STRUCTURE 73
6.7 ENROLMENT SERVER 73
6.7.1 DATA EXPORT, TRANSPORT AND MANAGEMENT 74
6.7.2 CIDR SANITY CHECKS 74
6.7.3 DATA ARCHIVAL 74
6.7.4 MAIN PROCESSING PIPELINE 75
6.7.5 BIOMETRIC DE-DUPLICATION 77
6.7.6 MANUAL ADJUDICATION 77
6.7.7 AADHAAR ISSUANCE & INTERNAL UPDATES 77
6.7.8 PRINT LETTER & LOGISTICS 78
6.7.9 TRACKING AADHAAR STATUS 78

6.7.10 E-AADHAAR 78
6.8 ENROLMENT MASTER DATA 79
6.8.1 INTERNAL MASTER DATA 79
6.8.2 EXTERNAL MASTER DATA 79

7 RESIDENT DATA UPDATES 81

7.1 RESIDENT DATA UPDATE 81
7.1.1 DEMOGRAPHIC DATA UPDATES 81
7.1.2 PHOTO UPDATES 82
7.1.3 BIOMETRIC DATA UPDATES 82
7.2 FINDING AADHAAR NUMBER 83
7.3 AADHAAR CANCELLATION 83
7.4 UPDATE CHANNELS 84

8 BIOMETRICS IN ENROLMENT 87

8.1 ACCURACY OF A BIOMETRIC IDENTIFICATION SYSTEM 87
8.2 DESIGNING FOR 1.2 BILLION RESIDENTS 88
8.3 MULTI-ABIS STRATEGY 90
8.4 ABIS API 91
8.5 HOW DOES A MULTI-ABIS SYSTEM WORK? 92
8.5.1 INSERT 92
8.5.2 IDENTIFY 93
8.5.3 PROBES 93
8.6 MANUAL ADJUDICATION 94
8.6.1 MANUAL DE-DUPLICATION STAGES 94
8.7 MEASURING SYSTEM ACCURACY 95
8.7.1 ABIS DUPLICATE "ANALYSIS" 95
8.8 CONTINUOUS MONITORING 96
8.9 BIOMETRIC SDK 97

9 ANALYTICS AND INFORMATION DISSEMINATION 99

9.1 INTRODUCTION 99
9.1.1 DATA FOR UIDAI'S INTERNAL CONSUMPTION 99
9.1.2 DATA FOR UIDAI'S ECOSYSTEM PARTNER CONSUMPTION 99
9.1.3 DATA FOR PUBLIC AT LARGE 100
9.2 METADATA STORAGE: THE ADW & DATAMARTS 100
9.3 SHARING METADATA: REPORTS, DATA SETS AND PORTALS 102

PART III: AUTHENTICATION & E-KYC 105

10 INTRODUCTION TO AUTHENTICATION 107

10.1 FEDERATED AUTHENTICATION MODEL 109
10.2 AUTHENTICATION API 110
10.3 BEST FINGER DETECTION (BFD) API 111
10.4 BIOMETRIC AUTHENTICATION 112
10.5 DEMOGRAPHIC AUTHENTICATION 113
10.6 ONE TIME PIN (OTP) AUTHENTICATION 113
10.7 AUTHENTICATION SERVER 114
10.8 INFORMATION PRIVACY & SECURITY 115

11 AUTHENTICATION ECOSYSTEM 117

11.1 UIDAI 117
11.2 AUTHENTICATION SERVICE AGENCY (ASA) 117
11.3 AUTHENTICATION USER AGENCY (AUA) 117
11.4 AUTHENTICATION MODELS 118
11.5 AUTHENTICATION DEVICE ECOSYSTEM 118

11.6 APPLICATION ECOSYSTEM 119
12 BIOMETRICS IN AUTHENTICATION 121
12.1 BIOMETRICS FOR AUTHENTICATION 121
13 ELECTRONIC KYC 123
13.1 E-KYC 123
13.2 SALIENT FEATURES OF E-KYC SERVICE 124
13.3 E-KYC ECOSYSTEM 125
13.4 E-KYC PROCESS 125
13.4.1 USE CASES 125
13.4.2 DATA FLOW 126
13.5 INFORMATION PRIVACY & SECURITY 126
PART IV: APPENDICES 127
14 LANGUAGE SUPPORT 129
15 REFERENCES 131

Table of Abbreviations

ABIS	Automated Biometric Identification System
API	Application Programming Interface
ASA	Authentication Service Agency
AUA	Authentication User Agency
BFD	Best Finger Detection
BI	Business Intelligence
BPL	Below Poverty Line
BSP	Biometric Service Provider
CIDR	Central Identity Data Repository
DDSVP	Demographic Data Standards & Verification Process
DoB	Date of Birth
EA	Enrolment Agency
ICDS	Integrated Child Development Services Scheme
JSY	Janani Suraksha Yojana
KYC	Know Your Customer
MNREGS	Mahatma Gandhi National Rural Employment Guarantee Scheme
OTP	One Time Pin
PDS	Public Distribution System
PID	Personal Identity Data
PII	Personal Identity Information (Personally Identifiable Information)
PoA	Proof of Address
PoI	Proof of Identity
RSBY	Rashtriya Swasthya Bima Yojna
SSA	Sarva Shiksha Abhiyan
UIDAI	Unique Identification Authority of India

Table of Figures

FIGURE 1 AADHAAR PARTNER ECOSYSTEM 25
FIGURE 2: AADHAAR SYSTEM OVERVIEW 33
FIGURE 3: SECURITY / PRIVACY FRAMEWORK 42
FIGURE 4ENROLMENT DATA FLOW – RESIDENT DATA TO AADHAAR LETTER 55
FIGURE 5ENROLMENT ECOSYSTEM 58
FIGURE 6: AADHAAR ENROLMENT SYSTEM 63
FIGURE 7 ENROLMENT CLIENT - BIOMETRIC DEVICE INTERACTION 66
FIGURE 8: BI TECHNOLOGY PLATFORM 102
FIGURE 9: AADHAAR AUTHENTICATION OVERVIEW 108
FIGURE 10 UIDAI AUTHENTICATION ECOSYSTEM 118
FIGURE 11: E-KYC DATA FLOW 126

Part I: Introduction & Overview

1 Introduction

This document outlines the UIDAI product strategy. This is a technical document, part of a collection of documents that are targeted at developers building open systems especially in the area of e-governance. This document must be read along with the following:

1. The UIDAI Approach Paper [Strategy, 2009]
2. Aadhaar Technology Strategy [Strategy, 2014]
3. Aadhaar Technology Architecture [Architecture, 2014]

In addition, readers are also encouraged to read the following documents for a better understanding of the Aadhaar system:

4. Notification creating the UIDAI [Notification, 2009]
5. The Demographic Data Standards and verification procedure Committee Report [DDSVP, 2009]
6. The Biometrics Standards Committee Report[Biometric, 2009]

While many elements of the system described by this document are in place, others are still evolving, and this document covers the thinking that guides their implementation. Further, detailed features may change over time, based on various policy considerations.

1.1 Primary Functions

Given the mandate of the UIDAI, the product must serve the following core needs:

1. Enrolment: Allow any resident of India to obtain a Unique Identification (Aadhaar) number.

2. Authentication & e-KYC: Authenticate the identity claim of Aadhaar holder and help Aadhaar holder use e-KYC instead of paper copies to facilitate electronic KYC verification.
3. Life Cycle: Allow residents to provide updates, and manage their identity data.

1.2 Product Goals & Principles

The Aadhaar is a program of the Government of India. It is not mandatory that residents enrol for this program. To make it attractive for them to enrol, this program must be adopted by other government programs, and private organizations to provide services and benefits to the residents. With these in mind, the goals for this product are:

- Social benefits
    - Reducing leakage in government social expenditure through removal of fake and duplicate entries in beneficiary lists by providing de-duplicated unique identity to residents.
    - Enabling inclusion, particularly financial inclusion by enabling opening of Aadhaar linked bank accounts and Aadhaar based financial transactions.
    - Enabling direct delivery of benefits to the resident.
- Technology benefits to the government sector
    - Stimulate adoption of technology and help drive e-Governance reengineering projects.
    - Increase adoption of open source and open architecture in e-Governance projects by demonstrating success within a large scale implementation.
- Benefits to organizations
    - Enable organizations to create a single customer master and provide a unified experience to their customers.
    - Help move away from paper based identity verification and KYC processes to a fully electronic platform, thereby significantly reducing operational costs.

1.3 Design Considerations

In addition to the goals, the product must be designed to mitigate adverse effects. The following sub-sections illustrate each of these considerations.

1.3.1 Design for Inclusion & Choice

The Aadhaar program is targeted to include those residents who are outside the formal systems, and may not have standardized identity documents. In fact, it is expected that Aadhaar will become the de facto identity document, and hence Aadhaar (and related standards) had to be designed for inclusion. While standardization of name, address, etc. are absolutely essential to create a common, electronically verifiable, national identity for all, it is also critical that these standards do not end up excluding sections of people. These aspects were considered by the DDSVP committee [DDSVP, 2009] to come up with verification procedures that do not exclude the marginalized.

Structure of the "Name" field of the resident is a classic example where a decision had to be made whether to use a western style naming convention where first name, middle name, and last name are captured separately or to use a single field name concept to accommodate all kinds of names from single word to multiple words. Considering the fact that, in India, several people have single word name and in some parts name can have more than 4 or 5 words, it was decided by the DDSVP committee to use a single field to capture name.

Similarly, in India, we neither have a standardized address format nor have well defined geographic boundaries beyond villages. This creates issues when trying to map addresses in a standard way. Various forms issued by existing registrars vary greatly when it comes to capturing addresses. It was felt that a standardized address structure must be created to ensure address is structured for electronic matching but at the same time caters to rural, semi-urban, and urban addresses.

It was also essential to ensure a full Date of Birth (DoB) with proof is not mandated since many people in India may either have no knowledge of exact date of birth or have no documentary evidence. To ensure no one is excluded due to this, the Aadhaar system allows either age to be captured or a full date of birth to be captured where available.

Another issue related to establishing identity is the fact that residents need to have some proof of their name and address while enrolling for Aadhaar. Although a large list

of Proof of Identity (PoI) and Proof of Address (PoA) documents are allowed, it was essential that the enrolment process also included residents who may not have any of the valid documents with them. To ensure inclusion, a concept of introducer was created that allowed residents with no PoI or PoA documents to be introduced by a valid listed introducer. In such cases, Aadhaar system captures the Aadhaar number of the introducer along with the enrolment. The Introducer concept was later extended to enrolment of family members through Head of Family (HoF).

In the absence of standardized primary identity documents, the Aadhaar is designed to include, but not depend on demographic data for the purpose of establishing unique identity via de-duplication.

The program was designed to account for resident choice – the first being that it is not mandatory for a resident to enrol in the program. Further, multiple registrars were selected – who may already be providing services to the resident, so that the resident could choose which one to approach for enrolment.

1.3.2 Ecosystem Approach

National identity projects are run in several countries by a single monolithic agency mandated to enrol its population. It was evident from the very beginning that given India's diversity – urban/rural, rich/poor, literate/illiterate etc. - the Aadhaar system would be best implemented by a set of cooperating partners or stakeholders.

Given the federal nature of the governance with strong state governments that implement many of the flagship schemes, it was important to enlist the state governments to enrol the residents of their respective states. The project needed enrolling agencies that would actually perform field enrolments on behalf of the registrars. These enrolling agencies also needed to procure standard enrolment kits. It was important to create a cadre of trained enrolment operators who work for enrolling agencies. This, in turn, required training and certification agencies. The ecosystem also needed device manufacturers and suppliers who would provide Aadhaar compliant devices for enrolment; this in turn needed a device testing and certification agency. A similar approach has been followed on the authentication and e-KYC services to ensure that the entire system can derive efficiencies from the Aadhaar system.

The UIDAI has also engaged in continuous dialog with standardization bodies, with technology providers, and other partners to ensure a high degree of uniformity in the quality of data collected and resident experience.

Such an ecosystem approach necessitated that the interfaces between these partners and systems were well defined and standardized. Hence, the UIDAI also needed to build a technology backbone that would hold together this partner ecosystem.

At the data centres where the enrolments were processed, UIDAI needed a set of suppliers to provide pieces of the Aadhaar backend system, most importantly, a set of Biometric Service Providers (BSPs) to provide multi-modal biometric de-duplication software solution that can de-duplicate the incoming enrolment requests and ensure that they are truly unique. In addition, UIDAI needed one or more agencies for application software development, 24x7 data centre operations, and security monitoring.

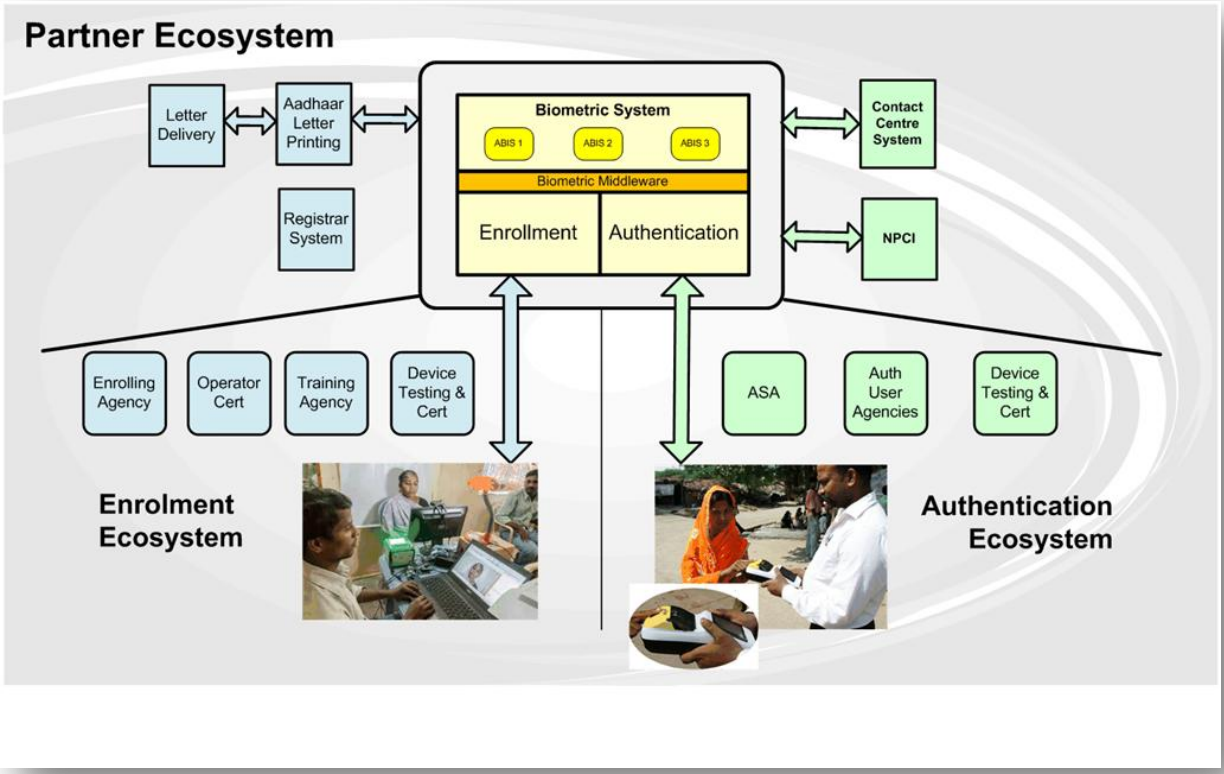


Figure 1 Aadhaar Partner Ecosystem

1.3.3 Identity as a Platform

One of the key considerations is to keep the Aadhaar system purely focused on identity and nothing else. The Aadhaar system only collects minimal data just enough to provide unique identity, issue the Aadhaar number after biometric de-duplication, manage lifecycle changes of that identity record, and provide an Application Programming Interface (API) for verifying the identity (online authentication) for various applications requiring identity verification.

Designing the Aadhaar system as pure identity platform allows clear separation of duties and leaves usage of identity to other partners, and their various applications which may be built on top of the Aadhaar platform.

1.3.4 Standardized Technology and Processes

The UIDAI has aggressively pursued the use of standards for interoperability, the creation of open interfaces, and the establishment of standardized, monitorable processes to ensure that residents get a consistent, high quality experience in their interactions with the UIDAI – while allowing partners the flexibility to innovate, and solve problems. This ranges from working with the national e-Governance standards for demographic data and structuring, standardization of the address structures, working with biometric device vendors for enrolment and authentication, etc.

1.3.5 Privacy by Design

Security and privacy of personal data has been fundamental in design of Aadhaar system without sacrificing utility of the national identity system. When creating a national identity system of this scale, it is imperative that privacy and security of personal data are not afterthoughts, but designed into the strategy of the system from day one.

Aadhaar system addresses these issues at its core. Following key aspects are results of this.

1.3.5.1 Aadhaar Numbering scheme

Aadhaar number is a random number with no built-in intelligence or profiling information. A 12-digit number was chosen based on the identification needs of the population in the next couple of centuries [Aadhaar Number, 2010].

It was also decided that, to ensure privacy, the date-of-birth and other attribute information should not be embedded in the number. Similarly, place of birth or residence using administrative boundaries (state/district/taluk) should also not be embedded within the number. When state/district IDs are embedded, the number faces the risk of becoming invalid and misleading the authenticator when people move from place to place. It can also lead to profiling/targeting based on the region/state/district that a person is from. The approach of storing intelligence in identification numbers was developed to make filing, manual search and book-keeping easier prior to the advent of computers. This is no longer necessary, since centralized database management systems can index the records for rapid search and access without having to section data by location or date of birth.

1.3.5.2 Minimal Data with No Linkage

Since Aadhaar system has data of all Aadhaar holders of the country in a central repository, it was essential that data be kept to a minimum so as to provide identity related functions (issuance and authentication) and nothing else. This design philosophy is derived directly from the fact that UIDAI respects privacy of the residents and not hold non-essential data within its systems.

In addition to having minimal data (4 attributes – name, address, gender, and date of birth - plus 2 optional data – mobile, email), this central database does not have any linkage to existing systems/applications that use Aadhaar.

This essentially creates a set of data islands containing resident data across various applications/systems (a federated model for resident data) rather than a centralized model eliminating the risk of a single system having complete knowledge of resident and his/her transaction history.

1.3.5.3 No Pooling of Data

Aadhaar system is not designed to collate and pool various data and hence does not become a single central data repository having all knowledge about residents. It has no

linkage information (such as PAN number, Driver's License Number, PDS card number, EPIC number, etc) to any other system. This design allows transaction data to reside in specific systems in a federated model. This approach allows resident information to stay in distributed fashion across many systems owned by different agencies. Any transfer or collation of data across these distributed information silos should follow due course of law.

1.3.5.4 Yes/No Answer for Authentication

Aadhaar authentication responds only with yes/no answer. For example, authentication answers questions such as "*resident claims his/her name is ..., is this correct?*" whereas it does not provide any scheme to ask questions such as "*what is the address of resident whose Aadhaar number is ...?*". Aadhaar authentication allows applications to "verify" the identity claim by the resident while servicing them while still protecting their data privacy.

1.3.5.5 Explicit Resident Consented e-KYC

A balance between 'privacy and purpose' is critical to ensure convenience of online identity is balanced with the requirement to protect resident identity data. External user agencies do not have access to the Aadhaar database. Aadhaar e-KYC service allows resident to authorize UIDAI to share electronic version of their Aadhaar letter. For every Aadhaar e-KYC request, only after successful resident authentication, demographic and photo data is shared in electronic format (via biometric/OTP authentication resident explicitly authorizes UIDAI to share electronic version of Aadhaar letter instead of sharing physical photocopies). Resident authorization is NOT used for multiple e-KYC transactions, instead, every time agencies require electronic version of Aadhaar letter data for KYC purposes, resident must authorize the agency.

1.3.5.6 No Transaction History

Aadhaar authentication does not have any knowledge of the transaction resident is performing. It only knows that resident is authenticated by this agency at this time. Aadhaar system has no knowledge and is not designed to keep track of specific transaction details such as depositing money or obtaining pension or anything else. This has been consciously designed to ensure resident transaction history is not part of a central system to ensure privacy of the resident.

1.3.6 Federated Data Model & One Way Linkage

While the UIDAI has the largest repository of identity data, one must resist the temptation to add all resident attributes to it. Each system that uses the data may have attributes that are important to their individual mission (for instance, the PDS requires information about income level, as well as family relationships), which may not be required by all systems.

It is also important that the various systems may have reference to the UIDAI (through the use of the Aadhaar number), but the UIDAI does not maintain a reverse links to any of these systems.

2 The Aadhaar System

2.1 The Aadhaar Approach to Identity

The UIDAI Strategy paper describes the need for a single, inclusive national identity system which reduces the costs of service delivery. Keeping in mind the learnings from previous efforts to issue national IDs, the UIDAI system was designed for the following characteristics:

1. Robust enough to eliminate duplicate and fake identities, and
2. Verifiable and Authenticated in a cost effective way.

This removes friction from various processes and brings down transaction costs, thus achieving the objectives of the Aadhaar platform.

2.2 Features of the UIDAI Model

The UIDAI model has the following features:

- Identity, not entitlement
- Identity, not citizenship
- A pro-poor approach
- Proper data verification
- Scaling via partnership model
- Flexibility for partners
- Enrolment not mandated
- UIDAI issues a number, not a card
- No Intelligence in the number
- Collect only basic information

- Process to ensure no duplicates using biometrics
- Online authentication & e-KYC
- Resident consented e-KYC
- Technology undergirds the system

These features are discussed in the Technology Strategy document [Strategy, 2014].

2.2.1 Demographic Data Collection

The UIDAI constituted a committee - the Demographic Data Standards and Verification Procedure (DDSVP) committee - with members from key government departments to identify the key data items that must be collected, and the process to verify them. The committee's report [DDSVP, 2009] forms the basis for the UIDAI enrolment process. The following fields are considered attributes of an individual's identity, and collected during enrolment:

- Name
- Date of Birth
- Gender
- Address
- For children below five, guardian's Aadhaar details
- Phone number (optional)
- Email address (optional)

2.2.2 Biometric Data Collection

The UIDAI also constituted a committee to define the biometrics that must be collected – to sufficiently identify individuals. The committee's report [Biometric Committee, 2009] was used to specify the biometrics that are collected during enrolment, as well as used for authentication. The following biometrics data is collected from the resident during enrolment:

- Facial photograph
- Fingerprints (all 10 fingers)
- Iris (both eyes)

Biometric modalities were selected keeping in mind:

a. **Standards**: Only biometrics that are mature, and have defined international standards should be considered. This ensures the presence of sufficient diversity in the number of devices, and technology providers, as well as the field research to support standardized processes for the collection of high quality data.

b. **Inclusion**: The collection of biometrics, and their use in de-duplication and authentication should not discriminate against people due to age, gender or profession.

c. **Differentiation**: Since de-duplication across the entire resident population is central to the uniqueness of Aadhaar, the biometrics (all put together) should have sufficient information to be able to de-duplicate across a 1.2 billion population.

2.3 The Aadhaar System

The following diagram provides a high level overview of the Aadhaar system.

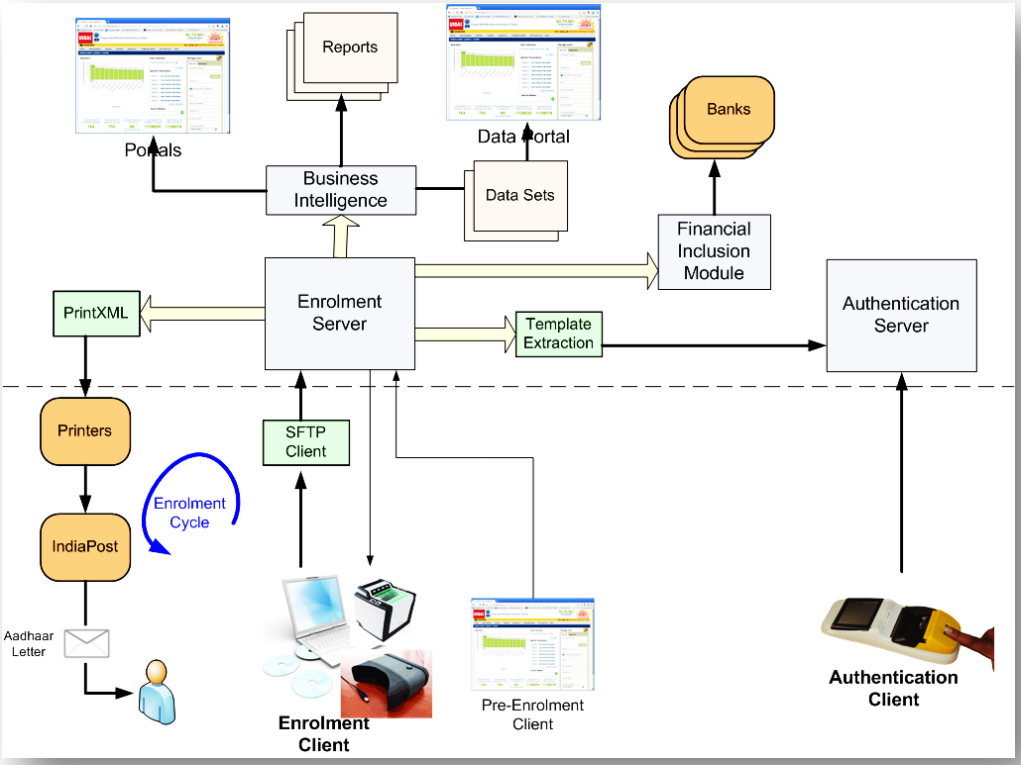


Figure 2: Aadhaar System Overview

2.4 Common Application Requirements

The UIDAI needs to support the diverse population of India, with varying languages, name and address conventions. Further, the population includes people in different occupations and age groups, some of which may result in fingerprints that are hard to capture. To facilitate this diverse population, all the applications within the Aadhaar system have common requirements, such as:

1. **Multilingual support**: Aadhaar applications support enrolment and authentication in all the official languages in India. Reflecting common practice, the enrolment is done in English and a local language. See Appendix for the list of supported languages.
2. **Common Master Data**: The UIDAI system uses various data as master data. This includes data that comes in from external bodies, as well as internally generated master data. The same master data is used across all applications. This includes:
   a. Location codes for States, Districts, Sub-Districts, Village / Town / City
   b. Postal Codes – Pincode
   c. Mapping between postal codes, and location codes
   d. Language codes
   e. Codes for internal entities – Registrars, Enrolment Agencies, AUAs, etc.

   This master data is used to control the quality of data entry, validate the data as it comes in. Changes to external master data have the potential to make current data invalid, and are managed through a supervised process.
3. **Common Name Structure**: The UIDAI follows the DDSVP committee recommendation that only a single, unstructured name be used.
4. **Common Address Structure**: The DDSVP committee recommended an address structure that includes a codified section (State, District, Sub-District, Village / Town / City, Pincode), a non-codified section (locality, street, house number), a landmark section, and possibly a care/of field. This structure is reflected in all public services of the UIDAI.
5. **Multi Modal Biometrics Support**: The biometrics committee recommended that the UIDAI capture 3 types of biometrics. All biometric sub-systems in the Aadhaar system support multimodal biometrics and rely on these for additional accuracy.

2.5 Enrolment

Enrolment is the process by which residents establish their identity, and apply for an Aadhaar. This process is carried out in the field through an enrolment ecosystem on behalf of the UIDAI, and supported through technology by the UIDAI.

The enrolment system is built on a robust platform, which supports all UIDAI processes. Since enrolment is an offline operation, and a long lived workflow, this platform provides UIDAI officials with the internal visibility to track enrolment packets (from enrolment till letter delivery) at an individual level. Additionally, analytics are available to provide insights at an aggregate level, which can be used to monitor the process, as well as identify improvements, etc. For instance, the Aadhaar system has visibility into the status of the packet at any time.

- On the enrolment client, through the use of the client synchronization with CIDR, the following events are captured, and sent to the server.
    - Packet type (New, Update, Correction)
    - Enrolment time
    - End of day processing – approval / rejection
    - Export / Re-Export
- On the enrolment server, following key events are available.
    - Receipt time at server
    - Quality check, and result of QC
    - Enrolment processing, and any checks that may have failed
    - Biometric de-duplication check
    - Aadhaar generation
    - Print Letter data generation
- On the logistics side, the following data is available.
    - Print letter data – sent to printer
    - Letter printed
    - Sent to India Post
    - e-Aadhaar

Further, any related correction / update requests (linked by enrolment id) can also be identified and tracked. This platform is also used to provide the same level of visibility to residents, and the partner ecosystem. Some examples of these services include:

1. Resident Portal: Provides information and services to residents such as status checks, e-Aadhaar, etc.
2. Partner portal: Provides partners with detailed information about their interaction with the UIDAI, including detailed reports, operational summaries, quality reports, etc.
3. Data Portal: Provides aggregated data for statistical analysis to the public.
4. Analytics portal: Provides a datamart for partners to perform analysis on aggregated data

2.6 Authentication

Authentication is the process by which residents assert their identity. The UIDAI provides an API, so that user agencies can integrate this into their process. Examples of this usage include authentication within a banking transaction, or establishing identity for the purpose of a government program like Public Distribution System.

2.7 Electronic KYC

The UIDAI ecosystem and the resident go through a process to establish the resident's identity which results in the issuance of an Aadhaar number. The rigor of this process is sufficient for certain service providers like telecom operators and banks to meet the Know Your Customer (KYC) norms established for their service. The UIDAI provides an API to registered entities to integrate an online KYC check in their process by providing electronic version of Aadhaar letter data with explicit resident authentication and consent. This is expected to simplify, and expand access to many services – like opening of basic banking accounts, issue of prepaid SIM cards, etc.

2.8 Platform and Common Modules

Core Aadhaar resident facing services (such as Authentication, Enrolment and e-KYC) are built on a robust platform. This includes internal components such as a workflow system to handle exception cases, fraud detection systems, as well as external systems such as an analytics portal.

The UIDAI platforms also support the non-functional requirements of the Aadhaar program, including performance, ability to scale out, etc.

2.9 Fraud Management

The Aadhaar enrolment and authentication systems are concerned with identity fraud. There are 2 primary systems concerned with fraud:

1. Internal analytics teams: Use various analytical tools and techniques to monitor enrolment and authentication processes, identify and flag outliers as suspicious activity for further investigation.
2. Fraud investigation workflows: Deal with internal, as well as external reports of potentially fraudulent activities. Such reports could be received from residents, or various UIDAI partners, and need to be investigated.

3 Information Privacy & Security

3.1 Introduction

The Aadhaar system is an identity system, dealing primarily with identity data, commonly described in literature as Personal Identifiable Information (PII). It is important that the system defines, for the residents, how this data will be collected, used, retained and protected from unauthorized access.

Aadhaar is a verified and reliable identity and hence is valuable for the purpose of transacting with the government, institutions, and adaptable for wider usage. This usage involves the processing of resident information, and transmission from user agencies to the CIDR. Such transmission and usage is covered by the information security policies.

Aadhaar has adopted the strategy of storing the minimal amount of identity data, that is required for its purpose, and in particular has chosen to stay away from maintaining transaction data regarding how the Aadhaar is being used (which is left to the user agencies). In this federated model, well defined rules and processes are required to protect collation of data across the entire ecosystem outside UIDAI based on larger privacy protection laws.

3.2 Data within the Aadhaar System

The Aadhaar system includes the following types of data

1) Aadhaar Number
2) Demographic Data
    a. Name
    b. Address (contains unstructured fields such as C/O, Street, Locality, and structured fields such as sub-district, district, state, & pincode)
    c. Gender (Male/Female/Transgender)

   d. Date of Birth
   e. Email Address (optional)
   f. Mobile Phone Number (optional)
3) Biometric Data
   a. Photograph (Face)
   b. Finger Prints (All 10 Fingers)
   c. Iris (Both eyes)
4) Enrolment records
5) Data change records
   a. Name change
   b. Address change
   c. Other Demographic Updates
   d. Biometric Updates
6) Authentication records
   a. Authentication User Agency
   b. Resident Aadhaar Number
   c. Time Stamp
   d. Authentication Factors Used
   e. Device Meta Data
   (No transaction data regarding purpose of authentication is stored)
7) External Master Data
   a. Administrative / Postal address hierarchy, with codes.
   b. Postal Code – mapping to addresses
   c. Language models.
8) Internal Master Data (Ecosystem Information)
   a. Registrars, and their authorized employees
   b. Enrolment Agencies, and their authorized employees
   c. Enrolment Station Operators, Supervisors
   d. Introducers
   e. Authentication Agencies
   f. Authentication Devices and related metadata
9) Aggregate Statistics

3.3 Personal Identifiable Information (PII)

PII refers to data which can be used to uniquely identify a single individual. From the data collected by the UIDAI the following is classified in this manner:

1) Aadhaar Number, along with other demographics data

2) Demographic Data (excluding Gender, Age / YOB, and structured components of address as long as they are not combined with other PII)
3) Biometric Data
4) Enrolment Records
5) Data Change Records
6) Authentication Records

While the demographic data that the UIDAI collects is already available with several agencies in the country – some of it is also available to the public (E.g., electoral rolls, railway reservation charts, telephone directories, etc.) - it is recognized as sensitive data within Aadhaar system, and handled with due care. Similarly, the authentication records are designed to be sensitive to resident privacy concerns.

3.4 Data Retention and Usage

The following table defines the period of time for which the UIDAI will store various types of information.

Data	**Retention period**	**Use**
Aadhaar	Forever	Used for authentication, future updates, etc.
Current Demographic Data	Forever	Used for authentication, future updates and communication with resident.
Current Biometric Data	Forever	Used for authentication, future updates.
Enrolment Records, and archived change records	Forever in archived form	Dispute resolution
Authentication Records	6 months in active audit and 7 years in archived storage to allow confirmation of a particular transaction only	Arbitration and dispute resolution. Users are alerted for each authentication transaction
Aggregate Records (Do not contain any PII)	Forever	Used to display statistics, and for other reports on the portal. Aggregate Records are

		freely accessible from the portal.
Master Data	Forever	Data validation

No PII data may ever leave the CIDR, except through an approved process, or with explicit resident consent.

3.5 System Security

System security is a significant topic that merits a separate document. However, some basic aspects are covered in this document to allow for the challenges to be appreciated, and designed against. The following diagram illustrates the resident touch points and security zones for the UIDAI.

Figure 3: Security / Privacy Framework

Application security for the above architecture cuts across all places where an untrusted source or destination is used. For example, the encrypted enrolment data file is uploaded in the DMZ (semi-secure zone outside production) to ensure it is scanned against Trojans or malwares before moving into production zone.

The enrolment/update data packets are encrypted by the client using public key cryptography with each data record having an HMAC which can identify any integrity violation of the data. Actual decryption of master keys are stored and managed within HSM (Hardware Security Module) appliance.

Following should be noted:

- Data about enrolment station, registrar, enrolment agency, operator, and supervisor is recorded and authenticated.
- Every packet is biometrically signed by operator (and supervisor in various cases) and contains complete process data including station id, timestamp, location (Pincode, GPS). This allows for strong validations and traceability.
- Every enrolment data packet is "always" stored in encrypted, tamper-evident files and are never decrypted or accessed during transit.
- Enrolment data is "never" decrypted until it is reached within UIDAI's data centre's secure production zone.

Usage of strong 2048-bit PKI encryption technologies ensures that no agencies or persons can access, modify, or misuse the resident data during field enrolment or in transit to the UIDAI data centres.

In addition to enrolment packet, resident data in Aadhaar master database and BI data store is protected through various security measures. These include:

- **Encryption** – Ensures data is encrypted and not available to administrators and other users in plain text format.
- **Anti-Tampering** – Ensures data is only altered by authorized applications and not via command line SQL scripts.
- **Data Partitioning** - Data is partitioned and stored across multiple databases with a random alias being the only link to ensure there is no central database containing all resident data.
- **Anonymization** – Hashing techniques are used for anonymizing data in BI/Reporting data store, and yet retaining the ability to do analytics.

Other than application techniques as described above to protect resident data, UIDAI has implemented data centre best practices and technologies such as firewalls, IPS

systems, zoning and access control, centralized security policy management, audits, 24x7 monitoring through Security Operations Centre (SoC), and strong security procedures used to ensure CIDR is protected.

The UIDAI remains cognizant of various threats to the security and integrity of the data in the system, as well as to privacy concerns of the residents. These are explicitly considered in the design and implementation of the entire system and its various sub-systems. The UIDAI has also published data protection and security guidelines for all the actors involved in the system, on UIDAI website.

4 System Interfaces

The Aadhaar system contains multiple interfaces between Aadhaar and external systems. This section documents the various interfaces and some key product principles. Additional details are available in various published documents.

4.1 Enrolment Client

The enrolment client is the first interface to the user. To ensure data privacy, security and to ensure standardization of the process, the UIDAI has decided to build and deploy this software. However, there are internal interfaces which are defined to ensure that the client and server communicate in a standard way. The enrolment client is responsible for ensuring that the processes specified by UIDAI are enforced on the field. Further, each client must be installed on a device, which is adequately registered and uniquely identifiable by the server.

The interface with the server includes:

1. The enrolment packet structure, including
    a. Encryption
    b. Biometric data structure
    c. Demographic data structure
    d. Audit Logs
2. The Enrolment ID structure
3. Client Sync
4. Data Export / Re-Export
5. Operator on-boarding and authentication

Other interfaces at the client include attached devices. These are specified below.

4.1.1 Aadhaar Biometric Capture Device

The enrolment client works with biometric devices from multiple vendors. A single interface has been defined to allow for simple plug and play from the field. This interface includes some degree of discoverability and the ability to identify device types as well as available functionality. The client modifies its behaviour appropriately (for instance, when the Iris device specifies that it can capture only 1 eye at a time, the client disables the dual eye capture).

The data captured from these devices is conformant to the relevant ISO specification (as specified by the UIDAI). The specification is published by the UIDAI [Device API, 2012], and compliance tested by the STQC.

4.1.2 Biometric SDK

The biometric data captured by the device is tested on the client for quality, presented to the user with appropriate feedback, and stored in an appropriate structure for transmission to the CIDR. This data is also tested for basic process compliance (for instance, uniqueness of all fingerprints captured, or matching against the operator's, etc.). All this requires a biometric SDK on the client. Multiple copies of this SDK, from different vendors, have been procured by the UIDAI, and are included in the client for this purpose.

The SDK specification is published by the UIDAI [Biometric SDK API, 2010].

4.1.3 GPS

The enrolment client and each enrolment are actively tracked using a GPS device attached to the enrolment station. This is to ensure process compliance and to provide visibility to the UIDAI as to where enrolments are happening. This interface is based on the NMEA-0183.

The specifications of the GPS device are published by the UIDAI, which empanels a set of devices for use in the field.

4.1.4 Document Scanner

The UIDAI enrolment process includes a scanner to scan documents on the field. This eliminates the requirement for a resident to bring a photocopy of the document (as it can now be considered 'Original Seen & Verified'). Further, this reduces the requirement of a form, and only one copy of the enrolment receipt needs to be generated. The resident and operator both sign the enrolment receipt which is scanned at the end of the operation and given to the resident. This ensures that the enrolment agency operations is completely paperless and secures resident data from leakage through this channel. The scanner is expected to comply with standard interface TWAIN.

4.1.5 Transliteration

A transliteration API has been created by the UIDAI, which allows real-time transliteration between English and the local language (bi-directional). This helps the operator to enter the data in one language, which is immediately transliterated (or translated, as appropriate). The operator can then correct the transliterated text (if required), before completing the enrolment. The final data can also be validated on the server using this API. In case, there is a significant difference, the data can be referred to a quality check operator to ensure that there is no significant difference between the two text components.

4.1.6 Master Data Import

The enrolment client depends on various master data definitions available from the enrolment server. This includes:

1. Address Structure data
    a. State, District, Village / Town / City names and codes
    b. Postal pincodes and corresponding Post Office names
2. Operator / Supervisor data
    a. Authorized operators for the enrolment agency
    b. Black lists for operators, and supervisors
3. Introducer data
    a. Authorized introducers

This data is imported on the client, from the server. It is important that this mechanism is secure to prevent errors and mischief (for instance, allowing unauthorized operators and supervisors to operate, or allowing incorrect data combinations to be entered on the field).

4.1.7 Secure Packet Transport, Upload & Management

The enrolment client exports data, which is then uploaded to the CIDR through a secure transfer. This involves the exchange of a list of files to be uploaded, and returns a status for each file. This includes the detection of errors (transmission / encryption), required resend of files, etc. In addition, it allows the upload of data only once – thus reducing load on the communication bandwidth and storage infrastructure.

A special client is available for this purpose to ensure that only authorized uploads are made to the server. This client uploads all of the exported data packets, the log files, and manifest to the server for processing.

4.2 Enrolment Server

The Aadhaar enrolment server receives enrolment data from the client, and processes it to make a decision on issuing an Aadhaar. There are different interfaces for this purpose. Some of the interfaces used by the client are also used on the server:

1. **Transliteration API**: Used to validate transliterations and identify operator overrides. These are manually checked to identify errors, and also to improve the transliteration engine.
2. **Biometric SDK**: Used to test data quality, and also match data against suspected duplicate enrolments.

In addition, the enrolment server uses the following interfaces.

4.2.1 ABIS

The ABIS API allows a biometric service provider to provide de-duplication functionality for the Enrolment server. This is used for new enrolments, as well as for biometric data updates. Multiple views of a resident's biometric details can be sent through the API. The API is specified as an XML over HTTP API, including a link to the

biometric data (which is fetched by the ABIS when required). Multiple, concurrent ABIS systems are used to enhance the performance, and accuracy of the system.

4.2.2 Logistics Partner Interface

Once an Aadhaar is generated for the resident, the data is printed and sent to the resident. This function is carried out by multiple logistics partners. A data specification is provided for this functionality, and a control interface is available to push files to the partner and receive status information from the partner.

4.3 Authentication API

Once Aadhaar generation is complete, the biometric data is sent to the Authentication server to provide authentication services for the resident. The authentication API provides services to various organizations that require authenticating a resident before providing a service. It could include just the Aadhaar or the biometrics along with the Aadhaar.

4.4 E-KYC API

This API provides services to those organizations which are required to meet certain KYC norms before accepting/servicing a customer. Organizations allowing use of Aadhaar to meet their KYC requirements may register with the UIDAI for this service. Every time e-KYC is done, the resident authenticates and authorizes the UIDAI to provide electronic version of Aadhaar letter to the service organization. The UIDAI in turn provides a digitally signed and encrypted copy of the demographic data and photograph to the service provider.

5 The Aadhaar Number

The UIDAI issues an identification number to all registered residents. This number is referred to as the Aadhaar Number.

5.1 Requirements

The Aadhaar number is designed with the following requirements.

- It must be large enough to accommodate population for the next 100-200 years.
- It must be small enough so that people can remember their own number.
- There must be a mechanism reserved to grow the number at a later point in time.
- It must not contain any intelligence, and hence cannot be used for any profiling.
- It must have inbuilt validation – allowing checking against typographical errors.

5.2 Solution

The UIDAI has implemented a 12 digit number that meets these requirements. These are detailed in a separate document [Aadhaar Number, 2010]. The salient points are:

- Numbers only – no alphabets, since this must be useful across all languages and easily usable on numeric phone keypads.
- First digit may not be 0 or 1. This is reserved for future extensibility.
- Last digit is a checksum digit – computed with the Verhoeff algorithm, protects against most simple errors like a single digit change, or transposing of digits.
- Length – 11 digits out of 12 (last digit being check digit) allows 80 billion possible numbers – enough to accommodate our population at least for next 100-200 years.

- It is possible to memorize a 12 digit number, since number does not change throughout the lifetime of the Aadhaar holder.
- Certain Aadhaar numbers are reserved for testing and promotions.

5.3 Format for Display and Printing

The number is written as 3 groups of 4 digit numbers, separated by a space. E.g. number shown below:

1234 5678 9012

5.4 Reserved Numbers

Certain Aadhaar numbers are reserved for testing, promotions, etc. These numbers may not be issued to residents. All numbers that match the following patterns may not be issued:

- Numbers starting with 0, and 1 – reserved for future expansion
- Sequentially increasing digits (cyclic), ex. 3456 7890 1234
- Six (6) repeated digits, ex. 9234 5692 3456
- Six digits reversed, ex. 9234 5665 4329
- Four (4) repeated digits, ex. 9234 9234 9234
- Three (3) repeated digits, ex. 9239 2392 3923
- Numbers starting with 9999 – used for testing

5.5 Random Number Generator

The Aadhaar system includes a random number generation sub-system which is used to generate random numbers, discard the ones that are not allowed, check for duplicates, etc. The random number generator has been created as a separate subsystem which can be scaled up independently, and supplies random numbers through a queue to the primary enrolment system.

Part II: Enrolment

6 Resident Enrolment

Enrolment is the process by which residents assert their identity and apply for an Aadhaar. This process is carried out in the field through an enrolment ecosystem and supported through technology by the UIDAI. This support takes the form of a standardized enrolment client, a standard process, an approved system configuration, certified devices, training material and certification process for operators, technical support for enrolment agencies, etc. The UIDAI also monitors the process for data quality and consistency.

6.1 The Aadhaar Enrolment Process

A resident approaches an enrolment centre to enrol for an Aadhaar number. Initially, enrolment is being done through enrolment camps organized for this purpose, and will eventually move to permanent enrolment / update centres.

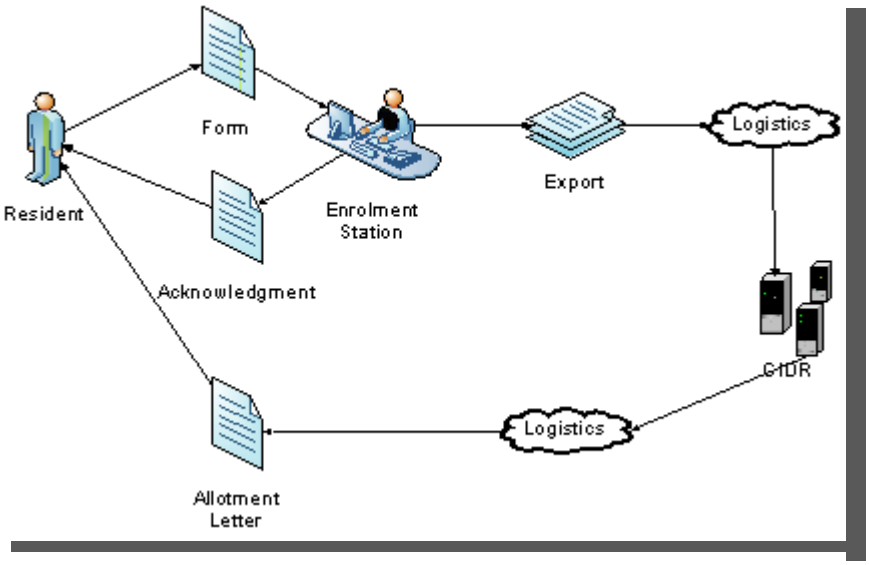


Figure 4Enrolment Data Flow – Resident data to Aadhaar Letter

At the enrolment centre, the resident is provided with the necessary forms that must be filled, and a list of documents required for validating the identity information. Resident demographic and biometrics data are captured at the enrolment station using standard Enrolment Client software provided by UIDAI. The identity of the resident and proofs are verified by a government approved verifier at the enrolment centre. The resident is issued an enrolment receipt. The resident data is then sent to the UIDAI for processing. On completion, the UIDAI issues a letter, which is delivered to the resident.

6.1.1 Resident Identity Verification

A resident is expected to provide documents supporting Proof of Identity (POI) and Proof of Address (POA). The list of documents was initially specified by the DDSVP committee [DDSVP, 2009], and has since been updated. See the UIDAI website for the updated list of documents. These documents must be produced in original form for identity / address verification at the centre. These documents are scanned and sent to the CIDR.

6.1.2 Demographic Quality Checks

The enrolment centre design requires a second terminal visible to the resident which mirrors the terminal seen by the operator – to enable the resident to verify the data as it is being entered. Further, the resident is advised to review the data at the end of the process. (Time spent on this screen is measured, and published in various operator wise reports and has been found to correlate with high enrolment quality).

The print receipt contains all the information. If a resident finds errors, the enrolment agency is required to go through a correction process immediately.

The operators and supervisors are also required to the check the quality of the demographic data locally, prior to exporting the data to the CIDR, and additional spot checks are carried out on the server. This has resulted in a higher quality of demographic data entry in the Aadhaar database.

6.1.3 Operator / Supervisor Verification

The enrolment process requires the operator (and supervisor in some instances) to biometrically certify that they have followed the process. This is done by on-boarding the operators to the local system, and performing a local biometric match during

enrolment. This ensures that each operator is valid, certified, and accountable for the data entered.

6.1.4 Exception Management

For various reasons, certain residents may not be able to go through the standard process of identity verification, and biometric data collection. To enable inclusion, and prevent people from being left out, an exception management process is defined, which provides for the following exceptions:

1. Biometric Exceptions – where the resident biometrics are not available, or of a poor quality that cannot be captured. This includes cases where specific biometrics are missing (like missing finger/eye).
2. Children below 5 years – where the biometrics are not considered stable. They are enrolled with a link to parent's/guardian's Aadhaar record.
3. Lack of identity documents – Residents who do not have such documents may rely on introduction from a government appointed introducer. The introducer is not required to be present at the time of resident enrolment, and may approve the enrolments later. The UIDAI has also provided for introduction by Head of the Family, who has all the necessary documents and has an Aadhaar.

An enrolment supervisor is required to validate all enrolments that include exceptions at the end of the day and approve them.

6.1.5 Field Monitoring and Quality Management

Process and Compliance Monitoring

The enrolment client, in addition to capturing the enrolment data and verifying the operator/supervisor, is also responsible to capture various metadata. This includes at least the following:

1. Station information
2. Peripherals connected – identifiers for all devices (where applicable)
3. Timestamps for various operations, including login, screen transitions, quality checks, biometric capture start and end, operator overrides, etc.
4. Biometric quality metrics
5. GPS data – place where enrolment is being carried out
6. Screen transition time and other usage metrics

This information is included in each enrolment packet and sent to the servers for analytics and reporting purposes.

Server Sync & Field operations monitoring

The enrolment station is also required to communicate with the UIDAI servers on a periodic basis. This allows the UIDAI to track the enrolment activity in the field, and also control various activities more effectively. Sync module checks the following:

1. Enrolments done on that station since last sync, packets sent to CIDR to identify data in transit, and list of missing enrolment files to resend. In case of a violation of enrolment agency SLA, this prevents further enrolments at this station.
2. Operator / supervisor black lists to ensure that they do not continue enrolments.
3. Enrolment client version, maximum packets accumulated since last export, etc. and if any UIDAI policy violation is detected, station is disabled from further enrolment until the issue is fixed.

6.2 The Enrolment Ecosystem

The enrolment ecosystem, shown here, contains multiple entities that work together to complete the enrolment process in a standardized, repeatable, measurable manner.

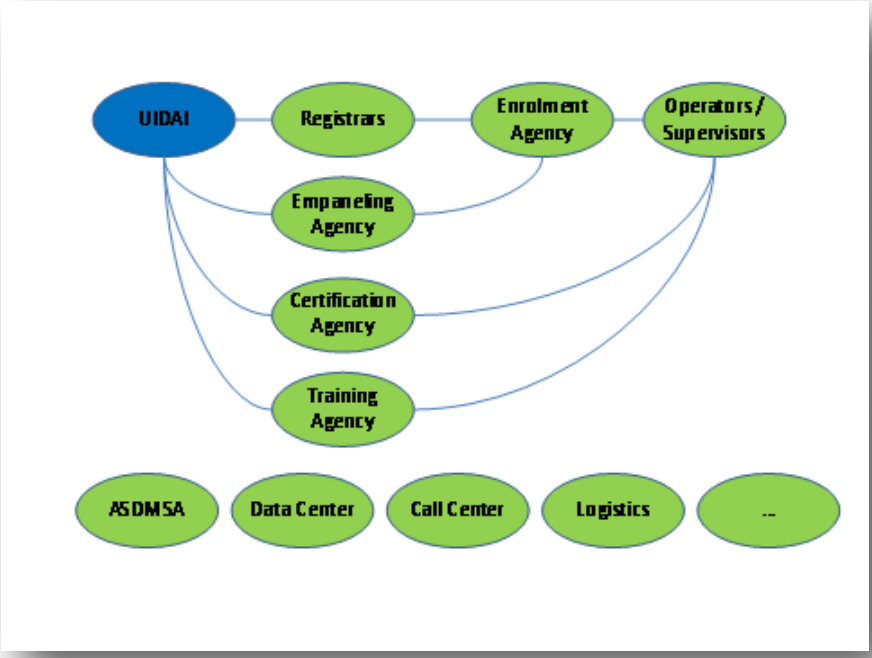


Figure 5Enrolment Ecosystem

The strategy document [Strategy, 2009] identifies the relationships, and roles of the following entities that are part of the core ecosystem. However, certain points are highlighted below.

6.2.1 UIDAI

The Unique Identification Authority of India is at the heart of this ecosystem, and is responsible for the definition of these relationships, and the core infrastructure. It is also responsible for measuring, and monitoring the performance of the system, and driving it towards delivering on its goals.

6.2.2 Registrars

Registrars are organizations that are already providing services to residents, and partner with the UIDAI in the collection of resident data. Since the registrars actively engage in the collection of data from the residents, they have access to the data from the point of collection. They may also collect additional data for their internal use.

6.2.3 Enrolment Agencies

These agencies are empanelled by the UIDAI and enter into a contract with a registrar to collect data. These agencies provide the field personnel, and equipment for enrolment. They are responsible for monitoring the field activities, adhering to field policies, conducting appropriate operator/supervisor training, ensuring that resident data is sent to CIDR on timely manner, etc.

6.2.4 Operators / Supervisors

These are employed by the enrolment agency to enrol the residents, and manage the enrolment centres. The operators perform the enrolment, while the supervisors manage the centre. The supervisors are responsible for the adherence to process, data quality, and exception management. Operators and supervisors must have Aadhaar numbers, and be certified before they can handle enrolment operations.

6.2.5 Training Development Agencies

The UIDAI has empanelled agencies to develop training materials for operators. These entities use the documentation, and new client releases to create training materials, including computer based training materials for operators. The training materials for each release are available to the enrolment agencies and others from the UIDAI website.

6.2.6 Operator Testing & Certification Agencies

The UIDAI has empanelled certain agencies to certify new operators. An operator who is trained, and has already been issued Aadhaar, can appear for the certification test. The certification agencies provide certification result to the UIDAI, enrolment agencies, as well as to the operator. Same is true in the case of supervisors as well.

6.2.7 Biometric Device Certification

STQC (Standardization Testing and Quality Certification Directorate), an attached office of the Department of Electronics and Information Technology (DeitY), Government of India, is the nodal agency appointed to carry out specifications as well as certification activity for enrolment and authentication devices requirements for the UIDAI. All the device specifications are hosted on STQC website and extensive certification activities are carried out at STQC labs at Mohali and New Delhi, which are equipped with state of art equipment to test and certify biometric devices. For more information, refer to STQC website [STQC, Certification, 2011].

6.3 Enrolment System Requirements

6.3.1 Language Support

The enrolment client allows data entry in English and one local language. The list of languages supported is derived from the Eighth Schedule of the Constitution of India. All Aadhaar services (enrolment, update, authentication, and e-KYC) support the use of resident demographic data in these languages.

6.3.2 External Artefacts & Documents for residents

In the enrolment process, various documents are provided to the resident. Each of these documents contains:

- Aadhaar logo
- Postal Address of UIDAI
- Web site Address and
- Call centre number

Each document (other than a form) contains a specific document number, which can be used to trace the document internally.

The following documents are a part of this process:

Enrolment Form

The UIDAI has a standard enrolment form, which includes all the information required from the resident. The form has space for information in 2 languages. The form is returned back to the resident after the enrolment process.

Enrolment Receipt

This document contains:

- All the resident details captured by the enrolment client in both languages – English and local language
- List of all the biometric data collected.
- Enrolment metadata – registrar, enrolment agency, operator, supervisor names.
- An enrolment number, which is unique.
- A URL to check enrolment status and Contact Centre phone number.
- Information about how to correct discrepancies – who to approach etc.

Aadhaar Approval Letter

The Aadhaar approval letter has the following fields:

- Name
- Relative Name (if provided)
- Date of Birth
- Address
- Photograph
- Aadhaar Number

- Information about how to correct discrepancies
- Contact information for UIDAI

The letter may also contain the resident's demographic data in the form of a machine readable code to improve ease of integration with other systems. The photograph is printed, so that even an illiterate individual can identify their own Aadhaar letter easily.

Similar documents are used for enrolment lifecycle-related transactions, such as correction/update of demographic/biometric data.

6.4 Device Ecosystem

The design of the biometric systems, including enrolment is discussed in detail in a later section. However, it is important to note that the UIDAI conducted extensive proof of concept studies in early 2010 [Enrolment PoC, 2010] for the enrolment process and carried out multiple workshops and consultations with experts from the field of biometrics, including device and technology vendors, and user agencies. These led to a good understanding of feasibility of enrolment using the prevailing biometric sensors (face, fingerprint, and iris).

Since no suitable standard interfaces were found for biometrics capture, the UIDAI developed the Aadhaar Biometric Capture Device API [Device API, 2010]. Enrolment device vendors are required to provide a vendor device manager, which complies with this API as a precursor to approaching STQC for certification. Aadhaar enrolment team provides testing support for VDM developers during and post certification.

In order to evolve enrolment station and devices layout in a typical enrolment centre, leading design consultants were consulted in addition to learning from POC. Design layouts were evolved both for the enrolment centre as well as for pack and move from one area to other. Device packing cases, where the complete enrolment kits can be packed and operated out of the box from carry case, were put together in order demonstrate the concept. Enrolment centre and enrolment station layouts were provided in the registrar on-boarding document in order to guide the registrars and enrolment agencies to set up their own enrolment centres.

The extensive engagement with user agencies, experts, and the industry has resulted in the creation of a level playing field, and a thriving ecosystem with a large number of

compatible devices. This has lowered the costs of the enrolment station and also contributed to the standardization of the enrolment experience.

A similar engagement exercise with user agencies, technology experts, and device vendors was conducted for authentication devices. This has led to the creation of a level playing field, with multiple partners providing certified solutions. Further, the use of POC studies, and data driven decisions has led to a standardized and high quality user experience combined with lower reject rates.

For more details, please refer to the UIDAI and STQC Websites.

6.5 Enrolment Module

The following diagram describes the Aadhaar Enrolment Application:

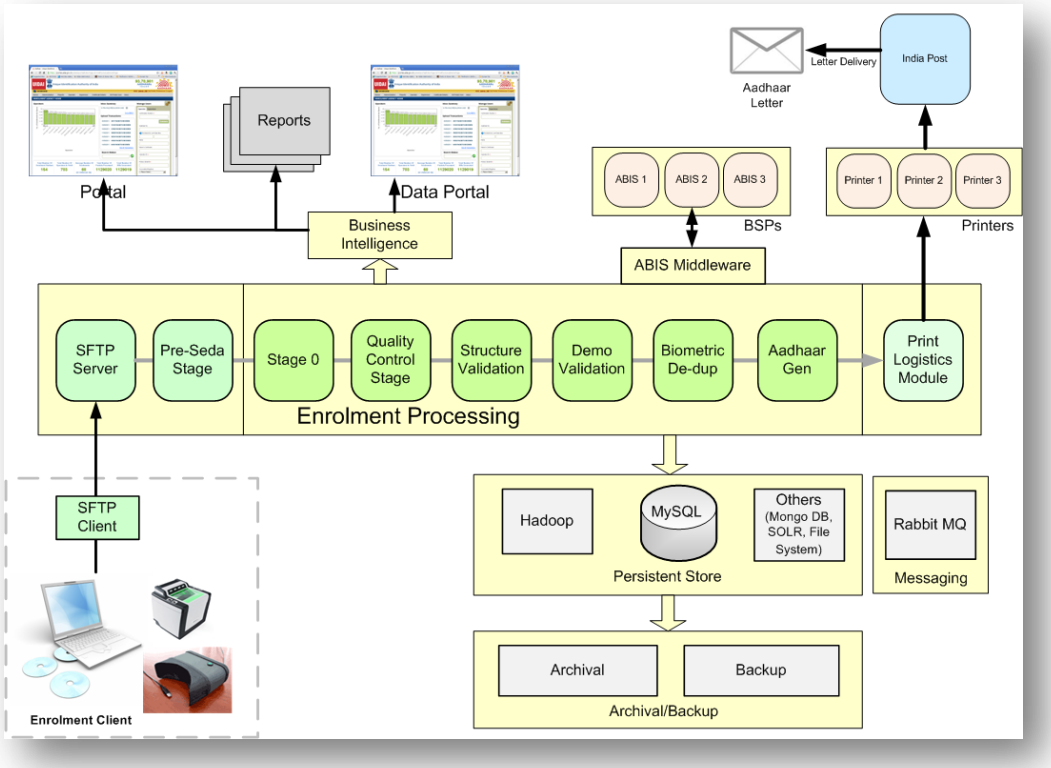


Figure 6: Aadhaar Enrolment System

6.6 Enrolment Client

The UIDAI developed the Aadhaar Enrolment client to support the enrolment process specified earlier, with adequate instrumentation to support all the monitoring and

management needs. This is available for enrolment stations (PCs based on Windows, and Linux, with supporting certified peripherals). The enrolment station provides a consistent, high quality experience to the users, and to the residents.

6.6.1 Client Features

The Enrolment Client (EC) supports the following features:

1. Enrolment
2. Corrections
3. Updates – demographic, biometric
4. System administration
    a. Operator management
    b. Quality Check
5. Client data sync
6. Enrolment data export

6.6.2 Device Support

The Enrolment Client (EC) supports all devices that meet the standards. Specifically, these include:

a. Biometric capture devices
    i. Fingerprint
    ii. Iris
    iii. Camera (face image)
b. GPS
c. Printer
d. Flatbed Document Scanner

6.6.3 Client Security

The enrolment client provided by the UIDAI has many layers of security and safeguards built into it to ensure that resident data is captured and transported to the server safely,

without rogue data getting into the system. The following are some of the checks that are built in:

1. **System Verification**: All enrolment stations are registered with the CIDR by enrolment agency personnel (different from operator / supervisor) authorized to do so. Each system is provided with a unique identity (not visible), which allows an enrolment to be traced back to a specific station, when required.
2. **Traceability to device**: Each enrolment data packet can be traced back to the enrolment station, and biometric capture devices used.
3. **System Sync**: All enrolment stations are required to sync periodically with the CIDR. Stations that fail to do so are barred from enrolment. This allows the UIDAI to identify rogue stations, and block them from enrolment. Further, enrolment metadata is available prior to the actual enrolment packets reaching the server – allowing an additional check against rogue data.
4. **Operator and Supervisor Verification**: All operators and supervisors in the system are registered, certified after testing, have an Aadhaar number, and are associated with enrolment agencies. Data quality is monitored for all operators, and any serious violations result in the immediate blacklisting of operators. The UIDAI also maintains a call centre to receive complaints from the field, which are investigated. These may also result in action against the operator.

   The enrolment client verified currently on boarded operators against the blacklist, as well as prevents blacklisted operators from getting on boarded.
5. **Operator / Supervisor Biometric verification**: All operators and supervisor operations are validated biometrically. At the time of on-boarding the operator, these are verified against the server. Once validated, during resident enrolments, the validation is local, and the data is bundled with the enrolment data and sent to the server. Thus, only authorized persons can perform enrolment, and they can be held accountable for their actions.
6. **Resident Data – Due Care**: During the enrolment process, all data captured from the user, including biometrics data, is maintained in memory and not written to disk. Only on completion of the enrolment process, is the data converted into the appropriate structure, encrypted with a 2048 bit public key and written to disk. This can only be decrypted at the CIDR. As a result raw biometrics and user data are never available in storage.

6.6.4 Biometrics Subsystem

The UIDAI uses biometrics to eliminate duplicates and ensure uniqueness during the enrolment process. The quality of the collected biometric data is critical for the accuracy of de-duplication and later authentication, and hence a key component to the success of the program. To achieve this while creating a standard enrolment experience, without limiting the choice of devices, the following steps were used:

1. Standardized devices – Certification process
2. Standard API to communicate with the devices
3. Standardized data collection process.- quality check and user feedback

6.6.4.1 Standardized Device & API Standards

All biometric devices used in enrolment are required to be certified by the STQC [STQC, Device Certification, 2010] for this purpose. This certification includes compliance with the Aadhaar Biometric Capture Device API, which is used by the enrolment client software to communicate with the devices.

The Aadhaar client interacts with the biometric devices through a two-layer structure, which is described in the figure.

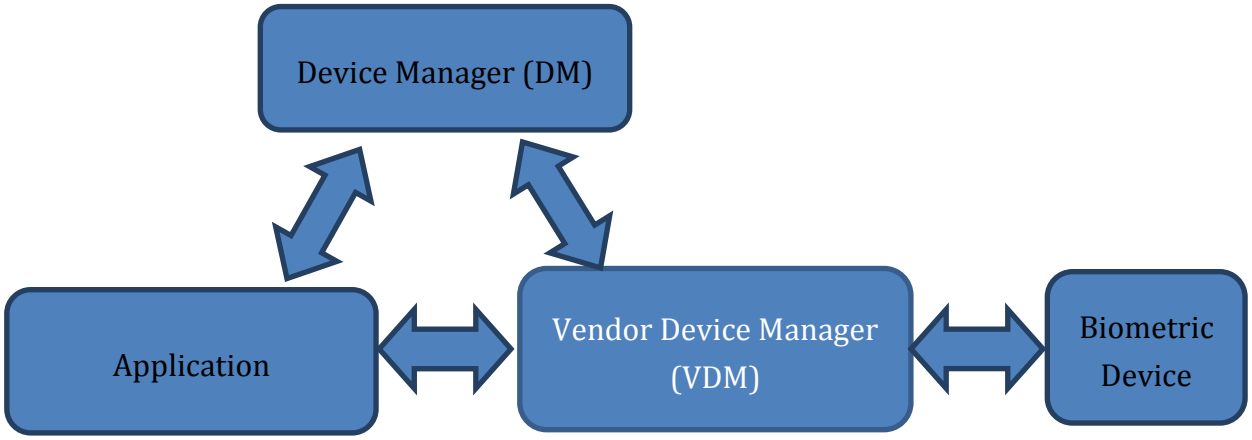


Figure 7 Enrolment Client - Biometric Device Interaction

1. **DM**: The vendor-independent Device Manager, which orchestrates the discovery of the VDMs by the application and manages connectivity to the VDM.

2. **VDM**: The Vendor Device Manager, provided by the device vendor, which manages the device and allows for biometric data capture.
3. **Application**: The Application that needs to use the biometric devices for capture. For instance, the Enrolment Client or the test harness.

The API is specified as communication protocol between the application, the DM, and the VDM. All communication is over TCP/IP sockets, which serves to isolate the software from multiple vendors into different processes and also provide platform independence. Applications could be developed in any language/environment such as Java, .NET, or C++.

The DM protocol allows device discovery, which ensures plug and play capability for devices from different vendors. Without any modification or settings in the enrolment client, during the field operations, enrolment agencies have the choice of devices to be used for biometric capture.

6.6.4.2 SDK based Quality Checking and User Feedback

The enrolment client enforces the data collection processes to ensure that correct data is captured. In addition, it has additional methods geared to reduce failure to capture, as well as to improve the quality of biometrics collected. For instance,

- **Auto Capture**: Each biometric capture device is required to have a built-in auto-capture capability which ensures that biometric images are captured only when deemed to be valid fingerprints slap or iris images and are of sufficient quality.
- **Forced Capture** is a technique to override the auto-capture capability. In this case the operator is responsible for the quality of captured image.
- **Multiple Captures**: Biometric Quality metrics are computed for all captures and presented to the operator. This is based on NFIQ score for the lowest quality finger. If the defined threshold value is not reached, the operator is required to recapture that image. For Iris, a proprietary quality score is used since there is no standard like NFIQ available.
- **Operator Feedback**: Actionable feedback is provided to the operator during biometric capture. Biometric data quality is measured using standardized automated algorithms and thresholds are utilized to decide whether a captured sample is of insufficient quality and warrants immediate re-capture. If the biometric capture is not of sufficient quality the resident would need to re-

capture the biometric. This is used to improve the quality of captures. This list includes:

- FINGER_MISMATCH: Mismatch between expected number of fingers and those captured.
- SLAP DRYNESS: Hands are too dry. Resident should moisten hands and try again.
- SLAP WETNESS: Hands are too moist. Resident should dry hands and try again.
- SLAP PRESS LIGHT: Hands not pressed properly against the scanner. Resident should try again with more pressure.
- SLAP PRESS HEAVY: Smudgy capture. Resident should try again, applying less pressure.
- IRIS FIDELITY: There is a focus or lighting problem, preventing an Iris from being recognized in the image. Attempt to resolve the issue after looking at the captured image and retry.
- PUPIL CONSTRICTION: The ratio of pupil to iris is too large. The resident should close eyes for a little while and retry.
- GAZE ALIGNMENT: The eye's gaze is in a direction different from the camera. Recapture after asking the resident to look directly at the camera.
- ILLUMINATION: A low illumination score indicates that the lighting conditions are too low, too bright, or not uniform. Repeat capture after fixing.
- EYE DISTANCE: The proportion of the distance between the eyes to the image size is incorrect. The distance between the camera and the resident needs to be adjusted.

Face image (photograph) quality feedback is based on a relaxed version of the ICAO standards.

- **Operator Overrides:** All operator overrides of policies are logged, and sent to the server, where these images are inspected for potential issues.
- **Best Capture**: If the capture meets the quality thresholds, it is sent to the server. In the event that it does not, then all capture attempts are sent to the server for processing.
- **Consistency Checks**: The client software also performs the following consistency checks to reduce process errors and handle exceptions.

- The client software confirms that all 10 captured fingerprints as well as the two irises are distinct. E.g., this would eliminate the capture of the left slap of the resident twice, instead of capturing the left and right slap once.
- The captured biometric is checked against that of the operator and the residents who enrolled previously on the same computer to avoid any chance of mix-ups. Any biometric capture either with the operator's or with a recent enrolment would be rejected and the biometric would have to be re-captured.
- Any biometric exceptions such as missing fingerprint or iris are logged and supervisor verification is required. In addition, an exception photograph is taken to facilitate manual verification.

6.6.5 Client Master Data

The enrolment client uses the following master data prior to getting registered:

1. Registrars
2. Enrolment Agencies
3. Mapping between Registrars and EAs

Once the client is registered for a particular Registrar/EA combination, the following master data is used for enrolments.

4. Valid operators and supervisors list for the EA
5. Blacklisted operators
6. Valid Introducers for the registrar
7. Address components master data
    a. States and Codes
    b. Districts and Codes
    c. Sub Districts and Codes
    d. Village / Towns / Cities and Codes
    e. Mapping from VTC to Pincodes
    f. Mapping from Pincode to Post Offices

It is important that the client continues to keep its master data in sync with the servers to ensure that all enrolments are valid. The following schedule indicates how this master data is updated and checked on the client. This validity is checked again on the server during each enrolment packet processing.

Master Data	Initial	Validity Checks	Updates
Registrars, EAs, and Mapping	Installation	At registration. On Sync (for current station).	On demand, from the portal
Location Master Data	Installation		On Demand
Operators / Supervisors	-	Operator On-boarding On Sync (for on boarded operators / supervisors)	On Demand (download from portal)
Blacklists	-	At operator login	Each Sync
Introducers	-		On Demand (download from portal)

6.6.6 Demographic Data Validation

The client includes multiple validation rules for the enrolment data. These are also re-applied on the server.

6.6.6.1 Name and Address Validation

All required fields, including the Name fields, are required to have no leading, and trailing spaces, and must have a length of at least 1 character. The address field should be structurally valid (i.e. the administrative hierarchy must be correct). Each of the fields has prescribed maximum length.

6.6.6.2 Transliteration

Every client system has the ability to capture resident data in 2 languages (English and a local language). The transliteration is based on a one to one correspondence at the word level, and is required to be exact. Since not all words are transliterated, it is important that the operator be allowed to override the transliteration. Transliteration support is available on a language basis from different providers.

The following requirements were used to evaluate, and select the transliteration providers:

- Transliteration in Unicode text.
- IME or Virtual keyboard as a widget to allow typing and correction synchronously.
- Keyboard and Mouse based navigation in case of Virtual keyboard.
- Reverse transliteration i.e. conversion from local language to English.
- Ability to predict multiple words, possible for the input provided in English.
- Ability to compare words of input and the destination language and return a score for the match.
- Usability aspects & intuitiveness of Virtual Keyboard or IME in case of changes and modification.
- Dictionary based lookup or logic based on Language input.
- Operating system support.
- Performance & quality issues based on the response time.
- Programming languages supported.

6.6.6.3 Pincode data & Administrative Regions

The UIDAI has obtained a mapping from every postal Pincode to the Village/Town/City that it serves. This unit (VTC) is available in an administrative hierarchy – VTC – Sub District – District – State. The enrolment client validates the hierarchy and forces the mapping between Pincode and VTC. This mapping is maintained through a portal, where enrolment agencies, registrars and UIDAI officials may add corrections, as well as fix the local language versions of these names. This master data is downloaded to the client on a regular basis.

6.6.7 Enrolment Id

The enrolment ID (EID) serves the following purposes:

1. Internal (UIDAI) – a way to identify each resident interaction – enrolment, correction, update, etc.
2. External (resident and ecosystem) – a unique token number for a resident to check enrolment status, lodge a complaint etc., till the Aadhaar number is generated. It is not required (and should not be used) for interactions after the generation of the Aadhaar and letter is delivered. This is also used by the

ecosystem for progress reporting and tracking enrolments as they make their way to the CIDR.

3. External (Registrar) – similar to the resident, except that the registrar receives a mapping from the EID to the Aadhaar number. This allows the registrars to correlate any data captured in the enrolment process with the eventual Aadhaar number.

The EID is generated for each enrolment, correction and update done on the client. It is also used internally for every update done on the server. The EID has 2 components:

1. Enrolment Number: This contains the Enrolment Agency ID, the Enrolment Station ID, and a running Sequence Number (limited to 5 digits, after which it rolls over to 0).
2. Enrolment Time Stamp in seconds.

This information is formatted for display, and data entry (in the portal, etc.) as:

xxxx/xxxxx/xxxxx dd/mm/yyyy hh:mm:ss, where ***x*** is a numeric digit **[0-9]**

Internally, this is stored as a 28 digit ID, formatted as

xxxxxxxxxxxxxxyyyymmddhhmmss

The enrolment packet file name contains the EID formatted internally as one of the components of the name. Initially, the enrolment file was just named <*EID*>.zip. However, this can be extended to include other information, such as

- Registrar Id
- Packet Type (Enrolment / Update / Correction)

External interfaces allow the user to get UIDAI online services with just the Enrolment Number or the entire EID.

The following checks are done on the client to ensure that the EID is unique:

1. Enrolment Station local time is checked against the server at registration. If the two differ beyond a set threshold, the station cannot be registered.
2. Enrolment Station local time is checked against the server at every sync call. If the two differ beyond a set threshold, the station sync cannot complete, and the station cannot continue enrolments.

3. The time stamp for an enrolment cannot be prior to the previous enrolment time stamp. If this fails, the system is locked till it can be synced.
4. There is no option to roll back the sequence number manually.
5. At registration, the sequence number is set by the server to ensure that it only moves ahead. (This is to account for systems deregistering and re-registering).

6.6.8 Resident Data Correction

The resident is provided with an enrolment receipt which contains all of the data provided to the operators. In case, the resident notices an error in the demographic data after this step, they can approach any enrolment agency and insist on getting the data corrected. A 72 hour window has been provided for this purpose, and the correction can be carried out on any enrolment station. Information about corrections is made available to the CIDR through the sync process.

6.6.9 Enrolment Packet Structure

It must be noted that the enrolment packet is constructed in memory and encrypted prior to writing of the file. All of the data, including biometrics and demographic data, is never stored in unencrypted form. The packet is encrypted with a randomly generated AES-256 symmetric session key and the key itself is encrypted with a 2048 bit public key, selected from a bank of UIDAI public keys.

6.7 Enrolment Server

The encrypted enrolment packets are shipped to the server via intermediate storage devices and secure transfer protocol to the CIDR. Each resident record (referred to as packet) goes through various stages before the final generation of the Aadhaar and communication to the resident.

6.7.1 Data Export, Transport and Management

6.7.1.1 Enrolment Client

The enrolment client provides a process to export the data to external media. This includes new enrolment packets and packets that are required to be re-sent. The client sends advance intimation about the enrolment packets through the periodic sync and receives an acknowledgement for packets received without errors at the CIDR. This allows a layer of protection against data loss in transmission. Further, since the process is done automatically, without human intervention, it reduces the management overhead of dealing with millions of files from all over the country.

6.7.1.2 Upload Client

The enrolment agencies collect exported data (always in encrypted form) from multiple enrolment stations and upload to the CIDR through a custom upload client. The client validates the credentials of the user performing the upload and ensures that only valid packets are uploaded to the server and only by authorized users.

6.7.2 CIDR Sanity Checks

The enrolment client is checked for validity – checksums, packet signatures, and with Anti-Virus software, etc. – in the CIDR DMZ before it is moved to the production zone of CIDR for processing.

Each enrolment packet is encrypted in the field with one of many public keys. The packet contains a key identifier (which allows the private key to be selected appropriately) and a checksum – this allows for the packet's integrity to be checked along the way. Packets that fail this integrity test are assumed to be corrupted en route (due to bit errors, or virus, system errors, etc.) and are required to be re-transmitted from the client.

6.7.3 Data Archival

Prior to being processed at the CIDR, the enrolment packets are archived to ensure that the data is kept securely. The archival system has the following requirements:

- All original packets (enrolments, updates, etc.) are required to be archived as-is, and "forever", ensuring high availability, and zero data loss.

- Data should be kept securely, and separated from core enrolment and authentication systems.
- Archival system may be physically disconnected (or at least parts of it should be); it should allow on-demand data retrieval with appropriate access control and approvals.

6.7.4 Main Processing Pipeline

After the sanity checks pass, the enrolment packet is passed onto the main processing pipeline. At a high level, this includes the following stages:

1. Automated Data Validation
2. Demographic De-duplication
3. Manual Quality Checks
4. Biometric De-duplication
5. Aadhaar issuance

Post Aadhaar issuance, the data is transferred to the logistics systems for delivery of information to the resident, and to the registrars.

6.7.4.1 Automated Data Validation

All validation checks that are done on the client for demographic data, are performed on the server again. These include:

- Name & Address validations
- Transliteration Validation
- Language Validations
- Pincode and Administrative regions
- Operator, Supervisor, Introducer Validations
- Other Data & Process Validations

6.7.4.2 Demographic De-duplication

Demographic de-duplication is used primarily to catch trivial duplicates(non-fraudulent cases where all the demographic fields are identical) that are inadvertently submitted to the system, e.g., when a resident has not received Aadhaar number in a few days and

decides to re-enrol at an enrolment station again. It is also used to de-duplicate children under the age of 5 year as biometrics data is not captured for children that young and residents for whom no biometrics data is captured (genuine biometric exception cases). The goal of demographic de-duplication is to filter these.

Demographic de-duplication within the Aadhaar context consists of the following steps.

- Finding suspected duplicates through demographic matching. A variety of matching strategies may be used to improve the demographic de-duplication to maximize the potential duplicates found at this stage. The fields searched during demographic de-duplication include:
    - Name
    - Age
    - C/O Name
    - Address fields (Building No, Locality, VTC, Pincode)
    - Mobile/Email
- Performing 1:1 biometric verification of the suspected duplicates found using demographic de-duplication to classify the suspected biometrics duplicates in one of the following categories:
    - Confirmed Duplicates
    - Confirmed Non-Duplicates
    - Requires Manual Verification

Demographic de-duplication is especially important when the percentage of duplicate enrolment increases for two reasons:

- A large number of duplicates can be identified in the demographic de-duplication stage and need to go through the biometric de-duplication which is computationally much more expensive.
- The overall system accuracy of the Aadhaar de-duplication system is improved since the demographic de-duplication filter catches a significant portion of the enrolled duplicates. This results in an overall reduction in the number of duplicates that go undetected since the accuracy of the biometric de-duplication stage is a percentage of duplicates that are sent to that stage.

Confirmed demographic duplicates could be used as probes to measure the accuracy of the ABIS systems.

6.7.4.3 Manual Quality Checks

Enrolment packets are sent for manual quality checks, where various quality check operators check the data for demographic, and image quality issues. This includes sanity tests against the resident image – existence of human image, gross errors in gender and age, as well as issues with the text data (transliteration errors).

6.7.5 Biometric De-duplication

Once a packet passes all validations, and demographic checks, it is sent to the biometrics sub-system for biometric de-duplication. ABIS systems from 3 different vendors are used to ensure the highest levels of accuracy and performance. The vendors are incentivized based on their accuracy and performance to ensure that they continue to improve the performance of their systems.

6.7.6 Manual Adjudication

All duplicates identified by ABIS systems are sent to the adjudication module. The purpose of this module is to ensure no resident is rejected due to potential occasional false matches of the ABIS systems. This module, in turn, uses an automated scheme to auto dispose sure duplicates based on additional SDK matching. For example, if ABIS-1 declares an applicant duplicate, the Aadhaar system has the capability to send the same request to the other two ABIS solutions and also use additional SDK based matching. This allows high percentage of duplicates to be disposed as rejects. For a small percentage of applicants with poor quality biometrics or matching scores that are in the grey area, a manual inspection is performed by adjudication team to validate. A maker-checker pattern is implemented to ensure “acceptance” of a system declared duplicate as genuine by adjudication operator requires additional supervisor level verification and approval. Daily performance reports are generated for system monitoring.

6.7.7 Aadhaar Issuance & Internal Updates

Aadhaar number is allotted on determining the uniqueness of the resident. The resident demographics data is associated with this Aadhaar number. This information is also

sent to the authentication systems, so that resident authentication can be performed successfully.

6.7.8 Print Letter & Logistics

Once an Aadhaar number is issued, and all internal systems are updated, the information is ready to be shared with the resident. This is done through various print and delivery (logistics) partners.

A print letter packet is created with all the information required to print an Aadhaar letter. Periodically, this information is grouped by language and postal pincode. It is further sorted by delivery post office and the village/town/city before it is sent to the print partner.

The print partner is responsible for printing the letter (including tracking information), and delivering it to the logistics partner. The logistics partner (India Post) is then responsible for the delivery of the physical letter to the resident.

6.7.9 Tracking Aadhaar Status

Residents can track the processing information of their enrolment packet through the resident portal. Portal offers a self-service option to track the enrolment using Enrolment Number or full EID. Tracking page shows the status of the packet within the enrolment flow and a message with appropriate action (if any) to be taken by the resident

6.7.10 E-Aadhaar

Residents can also choose to get an electronic copy of their Aadhaar letter through the e-Aadhaar service – where they can obtain a digitally signed PDF with the same contents as the Aadhaar letter.

6.8 Enrolment Master Data

6.8.1 Internal Master Data

The Aadhaar system contains a lot of internal master data, which is used to identify the various entities in the system. This includes the registrars, enrolment agencies, operators, and supervisors – as well as metadata about them. This also includes station level information – such as the unique station signature which is generated for each system by the client.

Additional master data is also available within the UIDAI for the purposes of financial inclusion. This includes all scheduled banks, their branch locations, and corresponding codes (IFSC), etc. In addition, the UIDAI has information about the lead bank for financial inclusion in a district. This information is used on the client for residents to provide some optional data to the UIDAI.

This master data is maintained, and managed by the UIDAI personnel, as well as the personnel from the various agencies through the UIDAI partner portal.

6.8.2 External Master Data

The Aadhaar system is also dependent on external master data which includes the following:

- **Administrative Regions** – Structure, Names, & Codes: This includes the states, districts, sub-districts hierarchy, as well as the individual Villages/Towns/Cities (VTC) in them. This information is obtained in a codified form from the Registrar General of India (RGI) and is further amended by updates from the field force. In particular, this data is also enriched with names in the local languages for the particular region.
- **Postal Data** – Post Offices, Pincodes, and regions served. This includes a list of all the post offices, their types, and the corresponding codes. This information is obtained from India Post.

The mapping between the post offices and the VTCs they service is also obtained from India Post.

All this information is enriched through input from the field through a crowd-sourcing approach. When an enrolment agency finds itself planning enrolments for a region, they are required to check that the master data for the region is valid. In case of discrepancies (spellings, missing or incorrect data), they make the corrections on an internal portal. This information is validated by the local post office and the master data is updated.

The enrolment agency can then download the updated data and enrol the residents.

7 Resident Data Updates

Since it is expected that the UIDAI will become the primary repository of identity data, it is critical to ensure that this data always stays current. The update module handles Aadhaar lifecycle update requests such as address update, mobile number update, photo update, etc. Aadhaar system allows residents to update demographic data and biometric data based on a set of published rules. All updates require resident authentication to ensure data of one resident is only accessible and updateable by that resident. The Aadhaar system currently provides both assisted and self-service modes of updates after due authentication of the resident and validation of appropriate documents. This module ensures all rules are validated, resident authentication is done, request origination is validated, and the after update, a notification is mandatorily sent to resident.

This is accomplished through the following processes.

7.1 Resident Data Update

Residents may update the demographic data with the UIDAI as it changes. However, there are reasons to update biometric data as well. This could be done to mitigate poor quality captures, or to accommodate changes in the individual with time.

7.1.1 Demographic Data Updates

The resident is allowed to update only fields that may have changed or were incorrect, and provide documentation for that request. All fields can be updated based on UIDAI policy. Date of birth field is restricted to 1-time correction (in case original data was incorrectly captured). The address is a single field, and must be updated in its entirety.

Various protections are built into the update process to prevent excessive updates to data that should not change frequently, such as Gender, Date of Birth and Name. The resident is provided with two options for the update process – either an in-person update at an enrolment station or a self-service update through an online portal (or mobile application in future). In all cases, residents are required to authenticate themselves (biometric or OTP), request the change, and provide necessary documentation to verify the data.

7.1.2 Photo Updates

Resident photo is integral part of Aadhaar. It is critical to be a recent photo of the resident since it is used across all identity verification systems. Although UIDAI, at this time, has not defined any mandatory photo update policy, it is highly recommended that residents update their photo at least once in 10-15 years. Photo update can be done at any time by the resident via a permanent update centre run by a registrar. In future, other channels such as self-service portal also may support photo update. Whenever the photo is updated, a new printed Aadhaar letter can be requested by resident. Both e-Aadhaar and e-KYC services automatically uses latest photo.

7.1.3 Biometric Data Updates

The primary use cases for biometric updates are:

1. Children reaching the age of 5 (and requiring first time biometric capture and full de-duplication).
2. Children reaching the age of 15.
3. Improving the quality of captures as compared to initial enrolment – this is recommended when the number of auth failures is high, but the fingerprints appear to be of good quality.
4. In the event resident biometrics undergoes any change over time the resident may update their biometrics.
5. Fixing incorrect biometric captures caused by field process errors. This rare case requires full biometric update and de-duplication.

The biometric data for the individual include the face photograph, the 10 fingerprints, and 2 iris images. To mitigate the risk of fraud, a resident is required to update the complete set together at the same time. Since all biometrics are required to be captured,

the feature is built into the enrolment client, and an enrolment station is used for this purpose.

On the server, the resident biometrics are compared against the previous biometrics, and replaced when the match is a high confidence match. In the event that the resident had no biometrics on record, the new biometric data is de-duplicated against the entire gallery, before the update is made.

7.2 Finding Aadhaar Number

It is common to misplace receipts, letters, etc. Until the usage of Aadhaar becomes common, it is quite possible that residents will want to find their Aadhaar number / enrolment status.

The Aadhaar system offers a mechanism for internal users (E.g., Contact Centre) to be able to lookup Aadhaar number based on data provided by the resident. Residents can call the Contact Centre for enrolment status, obtaining their Aadhaar number, or request for a re-print of their Aadhaar letter.

In extreme cases, where the resident is unable to provide sufficient data to uniquely find his/her Aadhaar record, it is still possible to find the unique record based on a complete de-duplication process via an equivalent of a full biometric re-enrolment process.

7.3 Aadhaar Cancellation

There are certain rare events which may result in the cancellation of a resident's Aadhaar. This includes the cases where the resident already has another Aadhaar number due to possible operator mistakes or has obtained the number fraudulently.

In all these cases, the Aadhaar number is cancelled by a designated UIDAI user after due internal approval process with audit trail. The number is never reused and once cancelled, all external service requests (e-Aadhaar, authentication, e-KYC, update requests, etc.) for that Aadhaar number will fail. Online authentication allows user agencies to instantly detect invalid Aadhaar numbers (potentially obtained fraudulently) before seeding them into their databases. However, the Aadhaar number and its history (of allocation and cancellation) will available to the UIDAI management for audit purposes.

7.4 Update Channels

To provide multiple options to resident to manage any data updates, UIDAI has created the following update channels:

1. **Permanent Update Centre**: Residents may go to a permanent update centre managed by a registrar where an operator helps resident with data process. There are two formats for the update centre – one allowing a complete data update including biometrics, and the second allowing only demographic data updates. During the complete update process, packets are securely generated and uploaded quite like initial enrolment.

   For demographics only updates, a lighter version of the enrolment client software is used. It is expected that the registrars will setup a larger number of demographic update centres to cater to the higher volume of these requests as compared to the lower volume of biometric update requests. In CIDR, these packets are scanned, validated, and processed via the update flow.

2. **Self-service Update**: Residents having registered mobile with Aadhaar system can go to online update portal managed by UIDAI and use OTP authentication to request data update. After the approval workflow, update packets are generated and passed on to automated update flow. This mode allows residents to request for any demographic data update. In future, newer self-service channels such as mobile application may be provided.

3. **Contact Centre and/or Postal**: A few limited fields are allowed to be corrected or updated by calling UIDAI Contact Centre and sending supporting documents via post. After manual inspection and approval, update packets are created by UIDAI and passed onto automated update flow.

4. **Update API**: People may change mobile numbers and address frequently as compared to other fields and it is critical that UIDAI provide maximum touch points for residents to be able to easily update their data. UIDAI intends to provide an "Update API" to select agencies (which are already authentication agencies) where resident can authenticate and request data updates.

Update features use common services that are part of enrolment server and most validations such as station validation, registrar/EA validation, operator/supervisor validation, etc. are identical to that of initial enrolment flow. The enrolment server performs a number of steps while updating/correcting the resident data.

An outline of the steps involved in the update flow is listed below:

- Receive update packets
- Validate and verify received packets
- Normalize demographic information contained in the packet
- Verify/update the biometric data contained in the packet
- Update the resident's data into CIDR system
- Manage printing and letter delivery.

8 Biometrics in Enrolment

8.1 Accuracy of a Biometric Identification System

A biometric identification system compares the biometric packet from each resident enrolment with all the existing biometrics in the gallery (1:N matching) to determine whether the resident has previously enrolled. If the biometrics does not match any of the biometrics in the gallery then the resident is determined as unique (and can be issued an Aadhaar number).

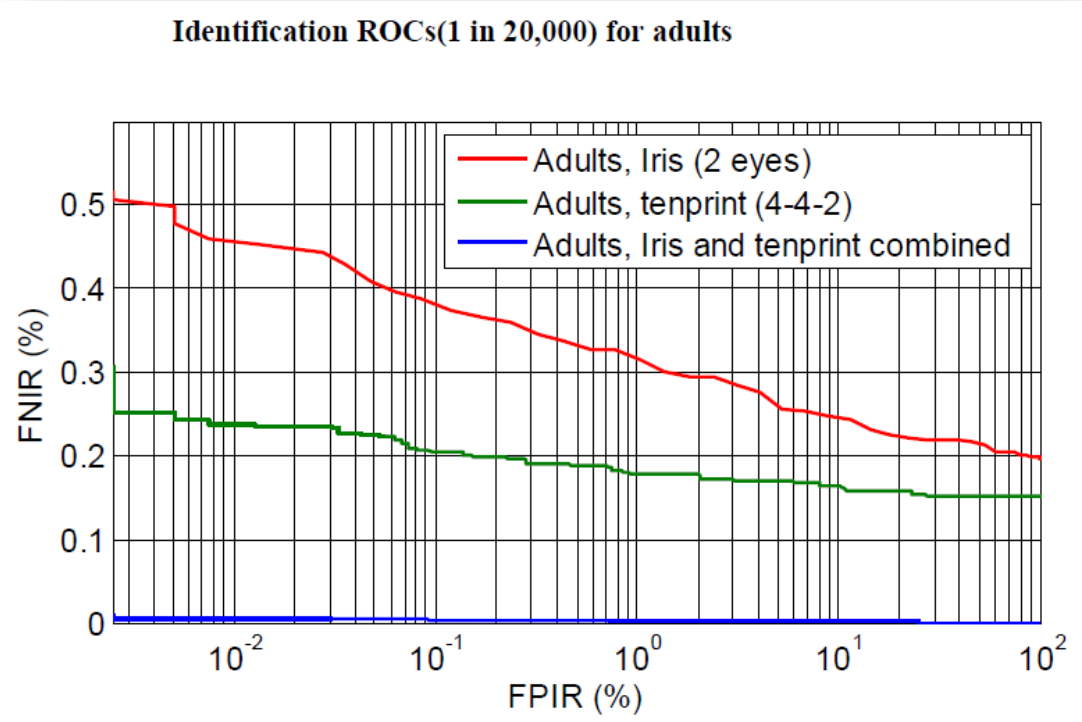
Biometric identification systems can fail in two ways:

- False Positive Identification: The biometric system incorrectly identifies a unique resident as a duplicate enrolment. This error can normally be mitigated through manual verification.
- False Negative Identification: The biometric system fails to detect a duplicate enrolment. This would lead to the issuance of a duplicate Aadhaar number.

The accuracy of a biometric identification system is characterized by the trade-off between the following metrics:

- **False Positive Identification Rate (FPIR):** The ratio of the number of false positive identifications to the number of unique residents enrolling in the system
- **False Negative Identification Rate (FNIR):** The ratio of the number of false negative identifications to the number of attempted duplicate enrolments into the system.

The trade-off between FPIR and FNIR of a biometric system is typically plotted on a ROC curve.



8.2 Designing for 1.2 Billion Residents

The Biometric Standards Committee was constituted to study the benchmarks for biometric de-duplication accuracy in other large systems around the world and recommend the approach to be followed by UIDAI to achieve a high de-duplication accuracy to scale to a population of 1.2 billion.

The committee report [Biometric Committee, 2009] found that while in the global context, 99% de-duplication accuracy had been achieved using fingerprints alone in a database size of 50 million, 95% de-duplication accuracy could be reasonably expected using 10 fingerprints and face at a database size of 1.2 billion in the Indian context. Hence, the committee recommended that UIDAI explore the use of the iris biometric in addition to fingerprint and face biometrics to achieve de-duplication accuracy in excess of 99% and also ensure total inclusion.

Based on the recommendation, the UIDAI commissioned a PoC study to determine whether multi-modal de-duplication could improve the de-duplication accuracy. The results of the enrolment PoC [Enrolment PoC, 2010] showed that an order of magnitude better accuracy could be achieved using a multi-modal de-duplication scheme including both fingerprints and iris.

This result was expected since the FPIR depends on the size of the gallery, while the FNIR is relatively independent of the gallery size. As the size of the gallery increases (as more enrolments are done) the ABIS solutions would need to tune their operating point on the ROC curve to match the FPIR requirements. For a large scale biometric system like Aadhaar it is critical to have an extremely flat ROC such that FNIR does not change significantly with the adjustment of FPIR. The inclusion of the iris modality resulted in a significant flattening of the ROC curve increasing the confidence that the system can scale to 1.2 billion people.

The PoC report also highlighted the increased inclusiveness of Aadhaar enrolment when iris was included. This is due to the fact that while fingerprint quality was variable among the rural poor due to occupations involving physical labour, the iris does not get worn out with age or use and remained unaffected by most eye surgery.

Finally the PoC report also observed that collecting and de-duplicating the biometrics of children is a challenge – face and finger biometrics are not stable until the age of 16. The iris presents a potential means to issue the majority of children a unique number linked to their biometrics, since the iris stabilizes at a very young age.

UIDAI decided to accept the recommendations of the PoC report and capture the following, biometrics for the enrolment of 1.2 billion residents:

- 10 finger prints
- 2 iris scans
- Photograph of face

The decision [Iris Inclusion, 2010] to include iris in the Aadhaar system was a considered one and took into account the critical needs of the project in ensuring the uniqueness of the Aadhaar number, and to also ensure that residents, particularly children and the elderly, are not excluded from enrolling for Aadhaar. The PoC empirically demonstrated that iris is easy to capture, highly accurate, and not too expensive.

8.3 Multi-ABIS strategy

The Aadhaar system is designed to service the entire population of India, and will involve the biometric identification of 1.2 billion residents. Since the estimated database size was an order of magnitude larger than the previous largest biometric database, the biometric subsystem needs to be constantly monitored for accuracy, scalability and performance.

Since de-duplication at this scale had not been previously attempted anywhere in the world, it was decided to procure 3 ABIS (Automatic Biometric Identification System) vendors to perform biometric de-duplication as a risk mitigation strategy.

Aadhaar is the first ever multi-ABIS system implemented in the world, and brings significant advantages:

- It ensures that the there is no vendor lock-in, if one of the ABIS vendors needs to be replaced (for whatever reason - technical or contractual) it can be done without bringing the entire system to a grinding halt.
- The three ABISs compete for work based on their accuracy and throughput. Since payment is based on successful de-duplication (not an upfront payment for software), solution providers compete to improve accuracy and throughput aiming for a higher volume of de-duplications.
- The deployment of multiple ABISs improves the accuracy of de-duplication. If any ABIS identifies a potential duplicate, it is sent to the other ABIS for verification. By combining the results of all 3 ABIS systems the overall biometric de-duplication accuracy goes up.
- The utilization of three different de-duplication engines with different implementations and different fusion strategies also helps to detect various kinds of software or data collection errors. In certain enrolments (for example in suspected duplicates and enrolments with poor quality biometrics) the enrollment data is sent to more than one ABIS to minimize the chance of an identification error.
- The use of multiple ABIS engines also permits continuous monitoring. Each duplicate found is used as “probe” to test the accuracy of the other two ABIS systems. This is critical to maintain the accuracy of the system on the ongoing basis.

8.4 ABIS API

To facilitate the concurrent operation of 3 ABIS systems and the future ability to integrate additional ABIS, each (ABIS) implements an interface that is compliant with the UIDAI specifications. That de-couples the biometric subsystem from the main application logic, and enables a management layer that can orchestrate across the multiple solution providers, continuously measure accuracy, performance and enable better decision making.

The enrolment server communicates with the ABIS over message queues. Each request is an XML string, and sent over the message queue. The ABIS is expected to send exactly one response for each request. The messaging subsystem is compatible with the AMQP protocol. There are 3 types of messages (request / response) per ABIS:

- Administrative Requests: All administrative requests are delivered on this message queue and the system is expected to respond immediately to these requests. Given the powerful requests that can be sent on this queue, additional security requirements are placed on this queue to ensure authorized use.
- Enrolment Requests: Most requests to the ABIS systems are of this type. These requests include operations to insert, delete, and de-duplicate the residents based on their biometric data.
- Authentication: "Verify" requests from external clients may be sent on this queue. Since authentication is an online process, the ABIS is expected to provide a rapid response to such requests.

Large data objects are not to be sent across these queues – instead a URL is provided in the interface to allow the ABIS to access the data. All biometric and demographic data is supplied to the ABIS through a URL. Dereferencing this URL provides a CBEFF (Common Biometric Exchange Format Framework) packet that contains:

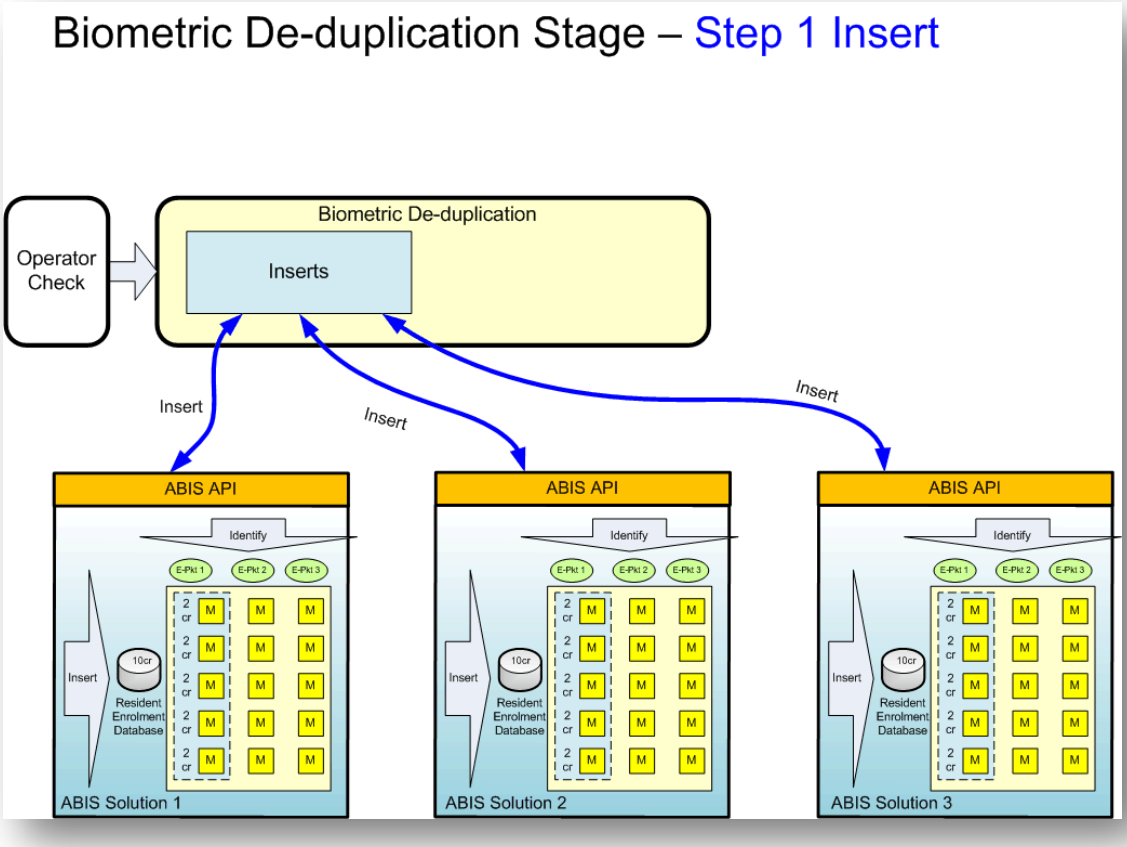
- Unsegmented fingerprint images
  - (4-4-2) [All UIDAI client data is captured in this format] or
  - (4-1-4-1), or
  - (2-2-2-2-2), or
- Segmented fingerprint images
- Iris images (Left, and Right),
- Face photograph

The CBEFF format provides multiple views for each biometric trait, though if the image capture is of sufficient quality only one image is provided per biometric trait. For Insert

requests, all images are constrained to specify the biometric positions that correspond to each image. For Insert requests, all images are constrained to be either uncompressed or lossless JPEG2000 images. The CBEFF packet is based on ISO/IEC 19785-3, with Patron ISO/IEC JTC 1 SC 37- Biometrics, Patron Identifier 257, Patron Format Identifier 7 – XML patron format. The details of the ABIS API can be accessed in the published specification [ABIS API, 2012].

8.5 How does a multi-ABIS System work?

The multi-ABIS system needs to maintain system synchronization among the ABIS providers. This is achieved through the following steps:



8.5.1 Insert

The biometric middleware sends each INSERT request to ALL 3 ABIS solutions with a URL to biometric packet to be inserted in the gallery. Before a de-duplication request can be sent for this biometric packet, ALL 3 ABIS must respond with an INSERT SUCCESS response for the INSERT request. **This is essential to ensure the correct**

operation of the multi-ABIS system. This implies that if one ABIS has a backlog in processing INSERT requests, it will impact the de-duplication **throughput of the entire system** since de-duplication requests cannot be sent. In such a condition an ABIS may be placed in INACTIVE state i.e. it cannot receive ANY de-duplication requests (to maintain correctness of system operation) until it has cleared the backlog of the INSERT requests. To rapidly clear the backlog of INSERT requests it is critical that the ABIS INSERT ONLY throughput is significantly higher than its de-duplication throughput.

8.5.2 Identify

Once all three ABIS have returned INSERT SUCCESS for a particular request, the de-duplication request can be sent. The de-duplication request is sent to one or more ABIS, determined by dynamic volume allocation. The dynamic volume allocation depends on measurement of the accuracy, throughput and hardware requirements of each ABIS solution.

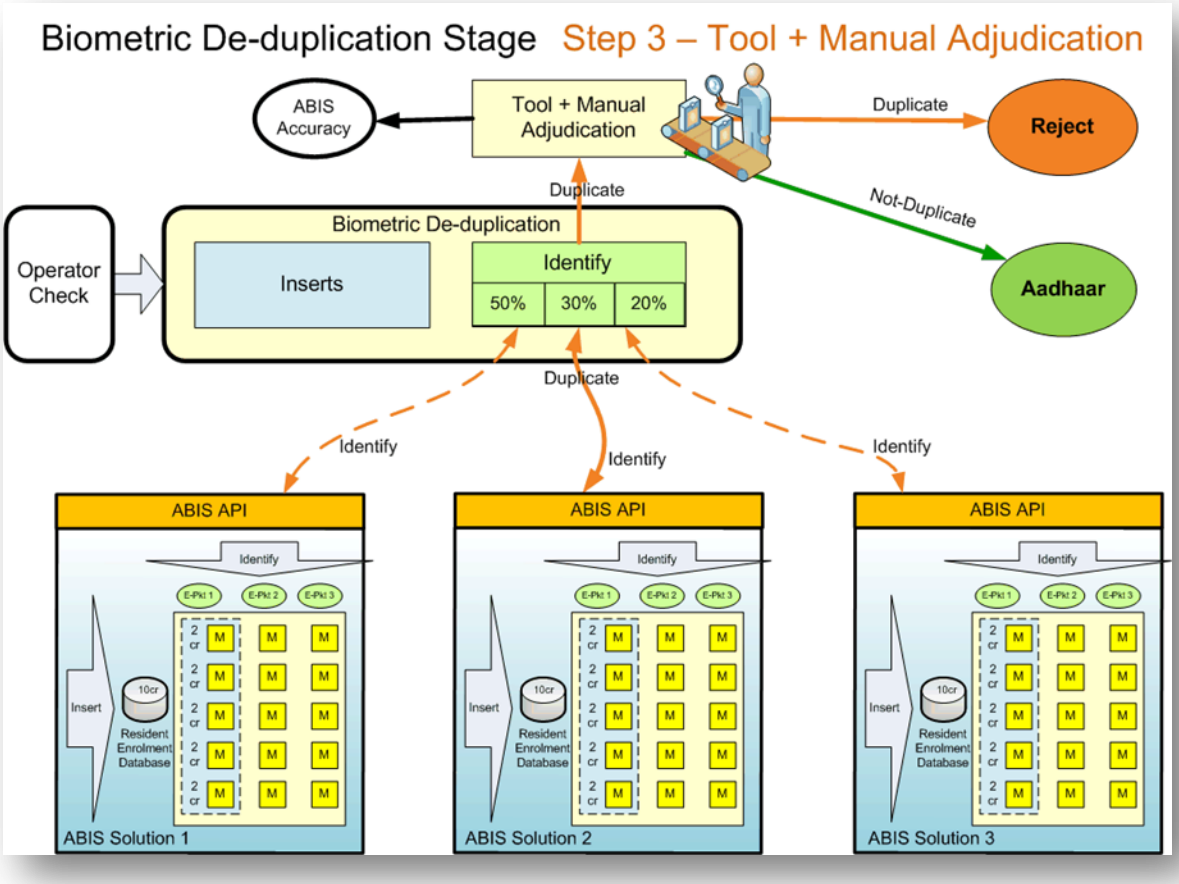
8.5.3 Probes

A probing strategy includes targeted gathering of data for use as probes, its entry into Aadhaar ecosystem and ABIS gallery, collection of responses from ABIS and its analysis to determine ABIS accuracy. Probes must test for matching data, as well as non-matching data. Certain probes may only use a single biometric, while others may use all the available data. The enrolment servers are expected to send probe requests to the various ABIS systems as a part of continuous monitoring, and testing. The probes are sent to all the ABISs, and the responses compared for accuracy. ABIS systems should not be able to distinguish between normal enrolment and probe packets. These PROBE requests are used to test the accuracy of the system and do not lead to Aadhaar generation.

Currently there are three types of probes:

- A percentage of de-duplication requests may be sent to two ABIS providers to check the accuracy of the de-duplication by comparing the results of the two ABIS providers.
- Known duplicates found using other methods such as demographic de-duplication are fed into the system from time to time to test the accuracy of the de-duplication system.

- A sample subset of duplicates found by one of the ABIS system is fed as a probe to the other ABIS systems.



The usage of probes allows for the continuous measurement of the de-duplication accuracy within the system.

8.6 Manual Adjudication

8.6.1 Manual De-duplication Stages

When an enrolment has been identified as a duplicate by one of the ABIS system, it then is sent to the Manual De-duplication (MDD) system. The MDD system has two main stages:

- Auto Disposition Stage: "Trivial" Duplicates are rejected during auto disposal stage. A set of rules has been established to confirm an identified duplicate without manual intervention. This has become important since the ABIS system

identifies up to 100,000 duplicates per day and it is not feasible to manually adjudicate all of them. The goal of the auto disposition stage is to eliminate about 95% of all duplicates identified by the ABIS.

- MDD Stage: Duplicates that could not be resolved (i.e. rejected) in the Auto Disposition stage reach the MDD stage. The interface at this stage contains all the demographic and biometric images from the resident enrolment and the "candidate" duplicate enrolments. This interface is for the use of trained biometrics operators and experts.

8.7 Measuring System Accuracy

8.7.1 ABIS Duplicate "Analysis"

All enrolment pairs that have been identified as duplicate by the ABIS system are analysed by the SDK using the "FP/FN" tool. Based on the analysis by the SDK each "duplicate" pair is classified into one of the following categories

- **True Duplicate**: Both resident and candidate enrolment are consistent. At least one modality clearly matches between the resident and candidate AND NO modality of good quality is a NON match. In this case the resident enrolment should be rejected, assuming that candidate already has an Aadhaar.
- **False Positive**: Both resident and candidate enrolment are consistent. No modality is a clear match between the resident and candidate. In this case the resident should be issued an Aadhaar especially if there seems to be no demographic match between the resident and the candidate.
- **Anomalous Biometrics**: Both the resident and candidate enrolment packets are consistent. At least one modality is both of good quality and NOT matching between the resident and candidate AND at least one modality is clearly matching between the resident and candidate. This indicates some error in the enrolment process and the resident needs to re-enrol. [In some anomalous cases it may be the "candidate" who needs to "update" their biometrics. It may be prudent to ask the resident re-enrol and the candidate to update biometrics].
- **Resident Inconsistent Biometrics**: Resident is deemed to have inconsistent biometrics if multiple attempts to capture the "same" biometric modality belong to two different residents. This happens when the biometric of another resident (typically operator/family member) get mixed up in the resident's biometric packet. This determination can be made ONLY if both attempts are of sufficient

quality and do not match each other. Enrolments having inconsistent biometrics are rejected and the resident is asked to re-enrol since they will not be able to use biometric authentication. It is expected that the number of residents with inconsistent biometrics will decline in the future for the following reasons.

  - The usage of consistency checking in the client.
  - The ability to delete a biometric capture in the enrolment client.
  - Sending only one image from within the enrolment packet if the image is of sufficient quality.

- **Candidate Inconsistent Biometrics**: Sometimes the duplicate "candidate" enrolment has inconsistent biometrics. In case, the candidates already have an Aadhaar, they need to update their biometrics. Otherwise they may not be able to authenticate.

Duplicate pairs determined either as "True Duplicates" or "False Positive" are used for the purpose of calculation of the FPIR/FNIR of each of the ABIS systems. Any true duplicate missed by an ABIS system is counted as a false negative, while any ABIS that finds a duplicate pair deemed to be false positive is counted as a false positive.

8.8 Continuous Monitoring

An important aspect of the continuous improvement process originates from the feedback that is provided from the server-side system to the enrolment agencies about the quality of their data. Monitoring and scoring the operators on various metrics is a key technique to improve the quality of the data.

- Poor quality of data capture in one or more biometric modalities: Some operators are poor at fingerprint capture, while others may be bad at iris. It may also indicate malfunctioning of the iris/fingerprint sensors.
- Number of anomalous duplicate enrolments done by an operator: A large number of anomalous duplicate enrolments is an indicator of process violations by the operator.
- Number of operator overrides or biometric exceptions: A large number of force captures and/or biometrics exceptions is an indication of process violations by the operator.

When enrolment agencies receive frequent reports on the quality of their enrolments, it leads them to improve their training and processes since their payments are linked to successful Aadhaar generated and not the number of enrolments conducted.

8.9 Biometric SDK

UIDAI has created a common biometric SDK API specification in Java and asked all the Biometric Service Providers (ABIS vendors) to implement this API. The main reasons to take such an approach are:

Vendor neutrality – Aadhaar system is implemented using open standards and standard APIs to ensure that all components across the system are neutral to proprietary and vendor specific features.

Interoperability – To allow various systems to interoperate in a seamless fashion it is critical that standard interfaces are used. This enabled common data format definitions and protocols across the components that expose similar functionality.

Use of best-of-breed algorithms – An open API allows best of breed algorithms to be used for special purposes. For example, if one fingerprint algorithm works well for old age people, and another one for younger people, a common API is required to dynamically choose and use one algorithm based on the input.

Plug-n-play capability – When multiple modalities and algorithms are used for true plug-n-play capability, common API and discovery mechanism is required.

Using this API, SDK developers may expose support for one or many modalities. For example, an SDK developer specializing in fingerprint algorithms may choose to implement only fingerprint modality support while some other SDK may provide support for fingerprint, face and iris modalities.

There are two major components exposed via this API from within the SDK:

1. Quality Check and Segmentation Engine: This interface is meant to expose quality check, segmentation, and sequencing functionality.
2. Extraction and Matching Engine: This interface is meant to expose extraction and matching functionality.

This biometric SDK has been used in a number of places in the Aadhaar ecosystem including:

1. Enrolment Client: Biometric quality and consistency check.
2. Authentication Server: Perform 1:1 biometric verification.
3. Demographic De-duplication: Perform 1:1biometric verification of suspected duplicates identified through demographic de-duplication.
4. Manual De-duplication: Perform 1:1 verification and quality check on per modality basis for each resident candidate pair identified by the ABIS system.
5. FPIR/FNIR Computation: Perform 1:1 verification and quality check on a per modality basis to measure the accuracy of the ABIS system.

For complete details about the Biometric SDK API, please refer to the published specification [Biometric SDK API, 2012].

9 Analytics and Information Dissemination

9.1 Introduction

Aadhaar application is designed to auto-capture metadata information which enables visibility into operations, and effects various improvements.

There are different stakeholders in the UIDAI project who have different uses for the different metadata which is captured. These users can be broadly categorized into three main groups, namely UIDAI-Internal, UIDAI-Ecosystem Partners& Public at large.

In the sections below we describe the kind of requirement that each specific data-user may have and how the UIDAI works towards fulfilling those requirements.

9.1.1 Data for UIDAI's Internal Consumption

The UIDAI internally requires data to understand the health of the various systems, throughput of various applications, quality and timing of field operations to detect and prevent fraudulent practices, compare operations of one geographic area with another, to compare one time-period of operation with another, to compare one partner operations with another, and share the best practices. In addition to regularly monitor the health of operations, applications, hardware, etc. there is a requirement to produce periodic reports & live dashboards for management monitoring.

9.1.2 Data for UIDAI's Ecosystem Partner Consumption

The entire UIDAI organization with its 8 Regional Offices, Head Office, a Technology Centre and two Data Centres is manned by around 400 officers (including contractual

consultants). With such a lean organizational structure, UIDAI is enabling operations of over 30,000 enrolment stations carrying out more than 1 million enrolments daily. This has been made possible by the creation of a thriving and active partner-ecosystem which implements the various operations. In such a set-up it is very important for the UIDAI to provide relevant information about the operations to its ecosystem partners, for them to monitor and improve their operations, which in turn benefits them and UIDAI.

To service the ecosystem partners they need to be provided with periodic reports and near-real time updates of their field operation status.

9.1.3 Data for Public at Large

UIDAI being a government organization funded by Central Government under the National Data Sharing & Accessibility Policy (NDSAP), is required to share anonymous datasets for public consumption especially researchers, about its operations. In addition to the policy requirement UIDAI also believes sharing anonymized data on its operations for data analytics by public at large may provide UIDAI with useful insights about its own operations which eventually may help UIDAI improve its operations.

To facilitate the requirements of the above three category of users, UIDAI has built systems which can provide end-to-end visibility of entire operations. In the case of enrolment, the metadata information starts becoming available to the UIDAI system from the time the enrolment is finished and the enrolment client is synched with the UIDAI system over the internet till the time the Aadhaar letter is delivered or the e-Aadhaar is generated and downloaded by the resident.

The end-to-end visibility of operations is enabled by a two-tier data warehousing structure comprising of the Atomic Data Warehouse (ADW) & Datamarts. The details on these are described in the subsequent section.

9.2 Metadata Storage: The ADW & Datamarts

The Atomic Data Warehouse (ADW) is the central repository of all processed packets of enrolment. All the information pertaining to the enrolment which was collected in the field is available in ADW. Examples include:

- Biometric Devices – Make, Model, Serial Number

- Screen Transitioning & Time spent on each screen
- People involved - Operator, Supervisor, End of Day Reviewer & Introducer
- Partners Involved – Registrar, Enrolment Agency
- Software – Client Version, Master Data Version
- Biometric Quality & No of attempts to capture
- Enrolment Start, End time, End of Day Review Time, Synch time, Upload time
- The time stamp of completion of packet processing
- The outcome of packet processing – Aadhaar Generation or Packet Rejection

The ADW contains information at a packet level and in extreme granularity. From a normal operations management perspective aggregations of this data are more useful. For example, UIDAI internally may want to know the distribution of the time taken by a particular version of the enrolment client software used by a particular set of operators or geography or age of resident, to understand the client software behaviour and consider improvement. Similarly, the Enrolment Agency having several thousand machines and operators deployed in field may be interested in observing the daily productivity or biometric quality of its machines operators and identify cases where it observes some anomaly which needs further attention.

To facilitate such requirements the datamarts are created. The datamarts are a collection of pre-defined metrics, based on current and anticipated operational requirement. These metrics are typically summation counts, averages of some operational parameter (e.g. average enrolment duration, average biometric quality, average number of slap scans for doing an enrolment, average number of enrolments done by an operator in a day, total enrolments with mobile number, etc.) which is captured under different dimensions (E.g. Registrar, Enrolment Agency, Resident Age Bracket, Enrolment Month etc.) of the operations.

The datamarts are loaded by periodically querying the Atomic Data Warehouse for the new packets that reach the ADW. The datamarts primarily serve two purposes:

1. It makes analysis of the meta-data which is there in the millions of packets created daily, tractable for meaningful insight and focused auditing of operations is possible; and
2. It significantly increase convenience and reduce time for extracting information for ready consumption.

To assist internal UIDAI users to leverage the collated data present in the datamart, UIDAI has deployed a point & click data extraction and visualization tool which creates queries for the slice and dice requirements. The datamart and the visualization tool together empower the UIDAI management to know a lot more about the operations in a structured manner and make informed decision making.

The logical view of the BI platform is given below.

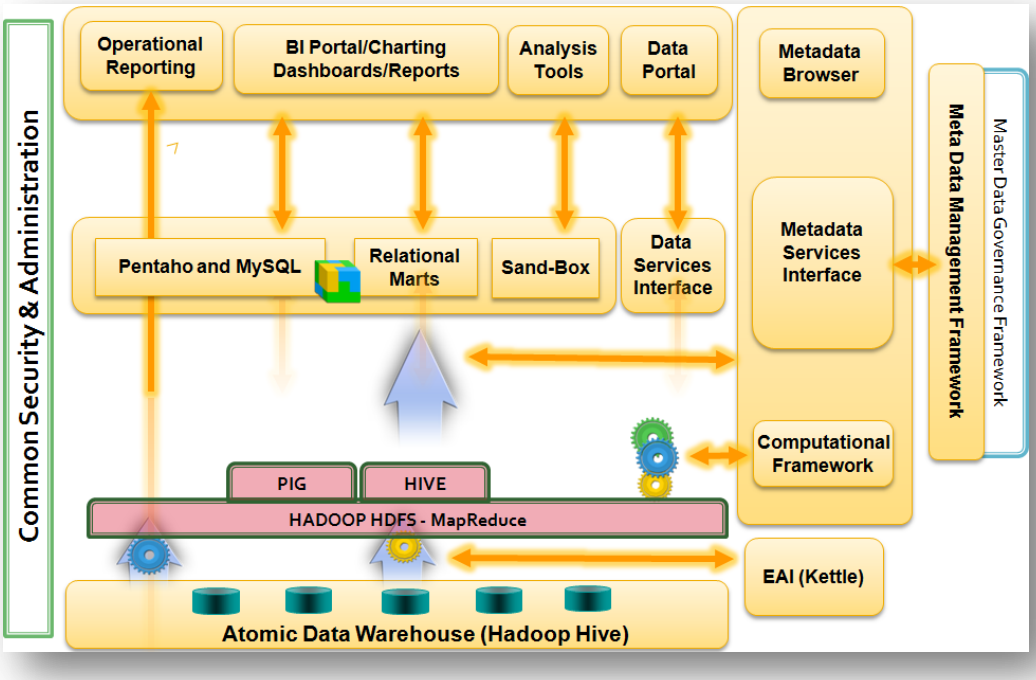


Figure 8: BI Technology Platform

9.3 Sharing Metadata: Reports, Data Sets and Portals

UIDAI shares the meta–data with its various users using a combination of reports, near-real-time portals and data-sets. The nature of the information decides the mode by which this information is shared. For example:

- The number & type of errors made by an operator is monitored over a period of time, say a month, and is made part of monthly operator monitoring report
- The number of packets which are ready to be uploaded (as known from synch information), yet have not been uploaded from a particular machine is part of a daily station monitoring report.

Information which is in some sense, always changing, and knowing its current status is of the most relevance is represented on a portal. For example:

- Number of stations of an enrolment agency active in the field in the last 24 hours.
- The number of Aadhaar generated till date by geographic regions.
- The number of packets uploaded in the last 24 hours for a particular registrar.

Datasets are used by UIDAI to share anonymized data pertaining to UIDAI operation with public at large. These datasets are available on the public data portal (data.uidai.gov.in). The datasets are generated daily in CSV format.

The current data sets are:

1. Aadhaar Generation by State & District – provides a geographical distribution of enrolments which have finished processing in the previous day.
2. Aadhaar Generation by Registrar & Enrolment Agency – provides a partner distribution of enrolments which have finished processing in the previous day.
3. Aadhaar Generation by Gender& Age – provides a demographic distribution of enrolments which have finished processing in the previous day.
4. Aadhaar processed in detail – provides a geographical, partner & demographic distribution of enrolments which have finished processing in the previous day.

There is a facility provided for downloading this data in an automated machine to machine format using a REST URL format and made available via data portal.

Part III: Authentication & E-KYC

10 Introduction to Authentication

While enrolment is the process where the UIDAI collects information from residents for the purpose of establishing a unique identity (the Aadhaar number), authentication is the process where the UIDAI allows any authorized agency to validate identity of individuals. This may be in the context of service delivery, or other use cases.

The Unique Identification (Aadhaar) Number, which identifies a resident, provides individuals the means to clearly establish their identity to public and private agencies across the country.

Three key characteristics of Aadhaar Number are:

1. Permanency (Aadhaar number remains same during lifetime of a resident)
2. Uniqueness (one resident has one ID and no two residents have same ID)
3. Global (same identifier can be used across applications and domains)

During the authentication process, the agency collects the Aadhaar number, along with other identity attributes (possibly including biometrics) and sends it to the CIDR for verification. The UIDAI responds with a Yes or No, thus authenticating the identity of the individual.

The authentication process is available, as an online API, allowing it to be integrated within any application. Since the response is immediate, it can be used in a wide variety of service delivery scenarios, where the identity of the service recipient must be verified.

At a high level, authentication can be 'Demographic Authentication' and/or 'Biometric Authentication'. In demographic authentication, the resident's demographic attributes are verified against the Aadhaar database. The resident may not be present during this interaction. During biometric authentication, the residents' biometric data is captured, and sent to the UIDAI for verification. The UIDAI mandates that applications should not use stored biometrics for this purpose, and hence a successful authentication implies that the resident was present during this process.

The following diagram depicts the high level overview of Aadhaar authentication.

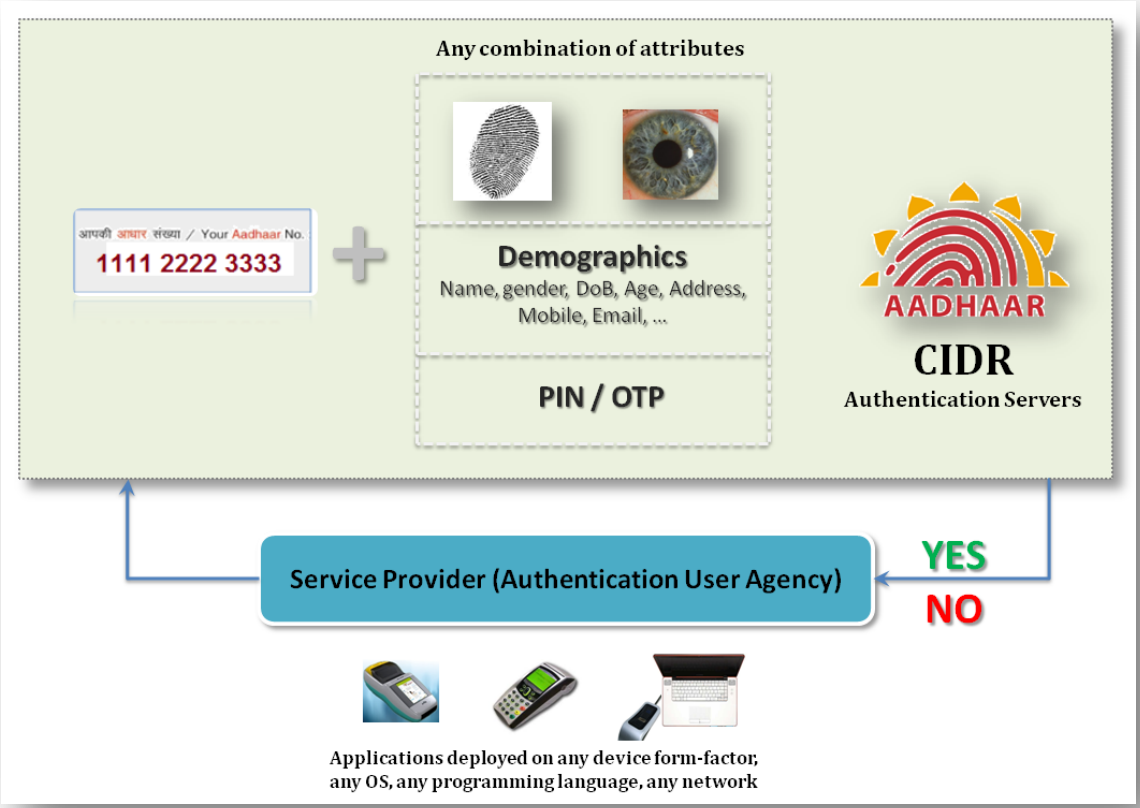


Figure 9: Aadhaar Authentication Overview

In general, following are the 3 categories of authentication factors:

1. ***What you have***: Something the user uniquely has (e.g., a card, security token, mobile phone, tablet/laptop computer accessing email, etc.).
2. ***What you know***: Something the user knows that is not public (e.g., a password, PIN, secret question, etc.). Demographic details such as date of birth may also be classified in this category although they are generally considered weak factors.

3. ***Who you are***: Something the user individually is or does (e.g., fingerprint pattern, iris pattern, signature, handwriting, etc.).

When user agencies adopt Aadhaar authentication, they can choose option 1 or 3 or both. Resident authentication can be strengthened by verifying the possession of the mobile by resident. One-Time-Pin (OTP) is a mechanism to do achieve this. Aadhaar authentication supports OTP and can be used by user agencies.

By combining biometric and demographic information stored with Aadhaar system, authentication of the resident could be strengthened. In addition, AUA specific factors such as cards may also be used in conjunction with Aadhaar authentication.

10.1 Federated Authentication Model

UIDAI offers Aadhaar authentication as a federated model. This implies that Aadhaar authentication usually works in conjunction with and strengthens an AUA's existing authentication system (as opposed to replacing an AUA's existing authentication system).

Most authentication systems could be described as "local" (i.e., pertaining to and/or valid for a few services, situations or entities) and "revocable" (wherein an existing identity factor could be revoked and reissued as a result of expiry, compromise or other valid reasons). Such revocable, local authentication systems come with set of strengths and limitations. Aadhaar authentication system, on the other hand, could be described as "global" (because of its applicability across situations, AUAs and services) and "non-revocable" (because Aadhaar identity factors such as fingerprints and iris scans cannot usually be revoked/replaced). Global, revocable/permanent authentication systems come with their own set of strength and limitations.

In the federated authentication model envisaged by UIDAI, the global-irrevocable authentication Aadhaar authentication system co-exists with and strengthens the local-revocable authentication systems of AUAs. It is expected that such a federated approach would eventually result in authentication systems which are stronger and more reliable than those that are based either only on global-irrevocable systems or only on local-revocable systems.

The following are some types of situations where an AUA could use Aadhaar authentication either on a federated authentication mode or on a standalone basis:

1. **One time usage**: When enrolling a new customer or creating a new service account for an individual. Examples are the issuance of a new PAN card, a new passport, creation of a new bank account or an internet service account for an online business. The AUAs in all such cases could authenticate an applicant's identity using the applicant's Aadhaar PID before issuing their own authentication factors.
2. **Periodic Usage**: AUAs can also use Aadhaar based federated authentication system for periodic update of their customers' (or employees' or associates') identity information. Examples are using Aadhaar authentication as a basis for renewing an Aadhaar holder's KYC data, the address of a bank account holder, etc.
3. **Transactional Use**: AUAs can also use Aadhaar based federated authentication system for carrying out any of their other business transactions. Examples include banks that authenticate a customer's Aadhaar PID as well as bank-related identity information (account number/user id along with password/OTP, etc.) before enabling banking transactions such as funds transfer, funds withdrawal, etc.

It is important however, to note that the federated model does not mandate the existence or use of an AUA's local-revocable authentication system in conjunction with Aadhaar authentication system. If an AUA so wishes, they could use only Aadhaar authentication to enable their services.

10.2 Authentication API

Unlike enrolment, where the UIDAI provides a complete enrolment client, authentication is a service provided through an API. This is to allow users to incorporate UIDAI authentication into their own application – either as primary factor, or as a secondary authentication mechanism.

The Authentication API requires that the agency collect user data, and send these as authentication factors to the UIDAI. Based on the needs of the service, different identifiers could be used along with Aadhaar Number. These identifiers could be combination of biometrics (such as fingerprints, iris impressions) and/or demographic

information (such as Name, Date of birth, Address) and/or a secret PIN or OTP number known only to the resident.

The authentication API is used by an application running on a host system which may be attached to a biometric sensor. The application is required to process the biometrics, and encrypt user's PII within the host application. The data is encrypted with the UIDAI public key, and hence cannot be intercepted en route to the UIDAI servers. This encrypted data is sent through the user agency's servers to the UIDAI. The user agency servers are unable to access the credentials, but may place certain policy requirements on it. This combined information (the user credentials and the agency policies) are signed by the agency and sent to the server.

The UIDAI server verifies the agency digital signature, decrypts the resident data, ensures that the policies are satisfied, and verifies the resident data against the information in its database, and responds with a Yes / No message, along with a signature that indicates the types of data used, and any error codes. The UIDAI response is signed with the UIDAI key, and can be verified through the standard PKI infrastructure.

If multiple authentication factors are used, each of them must individually successfully match with the data in the UIDAI for a "Yes" response.

The UIDAI has published the Authentication API on its portal [Authentication, 2012].

10.3 Best Finger Detection (BFD) API

During the initial authentication POC studies, it was identified that the quality of fingerprints may vary considerably between fingers of the same resident. This could be due to many reasons, including wear and tear on the finger, ability to present a finger to the sensor, or errors during enrolment.

It may thus be useful to appraise each resident of finger providing the best accuracy and successful matching results. We shall refer to this finger with best accuracy as the best finger. Resident may possess one or more best fingers. This knowledge allows the resident to provide his/her best finger(s) during authentication thereby increasing the chances of successful match.

Since a resident would normally not be aware of the best fingers to use for authentication, a Best Finger Detection process was defined where the resident would send a set of fingerprints (as captured from an authentication device) to the UIDAI server which would respond with the list of fingerprints that the resident could use for the purpose of UIDAI authentication. The value of this method was proved through additional POC studies that were carried out by the UIDAI [Authentication Accuracy Study, 2012], where the use of BFD brought down the False Reject Rate for residents significantly and thereby improved accuracy.

The UIDAI has provisioned the Best Finger Detection as an API, which is available to all user agencies [BFD, 2012]. UIDAI has mandated that this be used by user agencies to improve speed, and accuracy and overall customer satisfaction.

It is further recommended that residents be required to go through this process if they are unable to successfully authenticate themselves in a repeatable manner, and that all user agencies provision an appropriate application on authentication devices for this purpose.

10.4 Biometric Authentication

The UIDAI supports the use of multiple Who You Are factors. An application could select Fingerprints, and / or Iris for this purpose.

The following options are available to the application developers:

1. Fingerprints – 1 or more fingerprints could be captured from the resident. These could be sent in one of two formats:
    a. raw ISO images
    b. ISO minutiae

    The ISO minutiae take much less bandwidth, and are expected to be the primary format used in the field.
    The authentication API expects that if multiple fingers are used, they correspond to different fingers.
2. IRIS – 1 or 2 Iris images could be captured from the resident. Since there is no ISO minutiae format for Iris images, these are sent as images. However, the API does support different types of images, including KIND 1, 2, and 7 (also known as

"Uncropped VGA", "Cropped", and "Cropped and Masked") compliant to ISO 19794-6 standard.

The authentication API expects that if multiple IRIS images are used, they correspond to different eyes.

If multiple biometric types are used, they are treated as independent authentication factors. However, if multiple biometric captures of the same type (modality) are used, they can be fused and matched against the resident's data to improve the accuracy (reduced FRR) greatly.

10.5 Demographic Authentication

The UIDAI allows agencies to match demographic data collected from the resident against the corresponding data in the Aadhaar database. This allows for the use of Aadhaar as a proof of the identity, including all attributes (Name, DOB, Gender, Address), and corresponds to how identity matches are done.

All of these uses are possible with the Aadhaar authentication API, which allows matching of individual demographic attributes. The match criteria may include conditions, or allow approximate matches. This also allows additional checks against the identity of the users. For instance, a ticketing application may only want to know if a resident is above 60 years, to avail a discount, without necessarily wanting to know the date of birth of the resident. Similarly, a survey application may only need to validate the PINCODE of the resident, and not the entire data.

See Aadhaar Authentication API document for details.

10.6 One Time Pin (OTP) Authentication

Authentication focuses on matching a person's identity based on the reliability of a credential offered. Various agencies have different requirements for the degree of assurance required while authenticating beneficiaries/customers. When authenticating a resident, multiple factors may be used to strengthen the authenticity of the request. The choice of factors may also depend on the context of the authentication – E.g., it would be impossible for a website to authenticate a resident biometrically. Similarly, it would be impossible to use fingerprint authentication for a resident with no fingers.

In addition to the biometric factors, the UIDAI provides for authentication against an OTP, which can be delivered to the resident's registered mobile phone. This allows the use of a 'what you have' factor to replace the biometric.

An OTP request can be initiated by the resident by calling IVR or sending SMS or the request can be initiated by the application on behalf of the resident. Notice that OTP is always delivered on the resident's mobile/email and application is expected to capture that during authentication so that OTP can also be validated along with authentication. More details regarding the generation of OTP, as well as its use in authentication can be found in the Authentication API documents [Authentication, 2012].

10.7 Authentication Server

The authentication server is logically housed within the UIDAI CIDR infrastructure. However, it is a light weight server, easily replicated, and could be distributed for improved availability and performance. The server supports the entire set of Authentication APIs and methods. All requests and responses are logged for auditability.

The server is also architected to meet the following non-functional requirements:

- Multi-Data Centre architecture
- Sub Second response
- Fully load balanced
- Linear scalability with increase in number of servers
- Support for 10 Million Transactions per hour at launch
- Guaranteed Audits of response
- Updates for user preferences, OTP requests
- Demographic data updates from enrolment / update processing

10.8 Information Privacy & Security

The UIDAI Authentication APIs provide the ability to validate identity. In the process, user data is sent from the user agency to the UIDAI. The following points are relevant in this context:

- A 1:1 match is performed on the basis of the Aadhaar Number.
- No user data is returned from the APIs, only a Yes / No is returned.
- All user data, including biometrics, and demographic data, is encrypted from the host device in the field, and cannot be seen by any application en route to the server.
- Only authorized user agencies can access this API. They are expected to digitally sign all requests.
- Only authorized server agencies may connect to the CIDR for authentication purposes. This prevents the service from being exposed to DOS and other attacks.
- The UIDAI response is signed digitally, thus assuring the user agency that the authentication response is indeed from the UIDAI.
- The resident is alerted for all resident-present transactions (configurable through resident preferences).

These measures protect resident data, and allow the use of UIDAI authentication to make transactions non-repudiable.

11 Authentication Ecosystem

Similar to the ecosystem created for Enrolment, the UIDAI has created a network of partners to allow UIDAI authentication to be widely available, while still maintaining a great degree of accountability in the system.

11.1 UIDAI

The UIDAI is the central entity involved in the UIDAI authentication process. It is the holder of the identity data, and oversees the entire process.

11.2 Authentication Service Agency (ASA)

ASAs are entities that provide network services. They are required to have secure leased line (or MPLS) connectivity to the CIDR. They are required to ensure that only requests from valid user agencies are transmitted to the CIDR.

11.3 Authentication User Agency (AUA)

An AUA is any entity that uses Aadhaar authentication to enable its services and connects to the CIDR through 1 or more ASAs. An AUA enters into contractual relationships with the UIDAI, as well as the ASAs.

The AUA is responsible for the various devices that connect to the UIDAI system. It is responsible to ensure that these devices are certified by the STQC, and that all processes prescribed by the UIDAI are followed.

11.4 Authentication Models

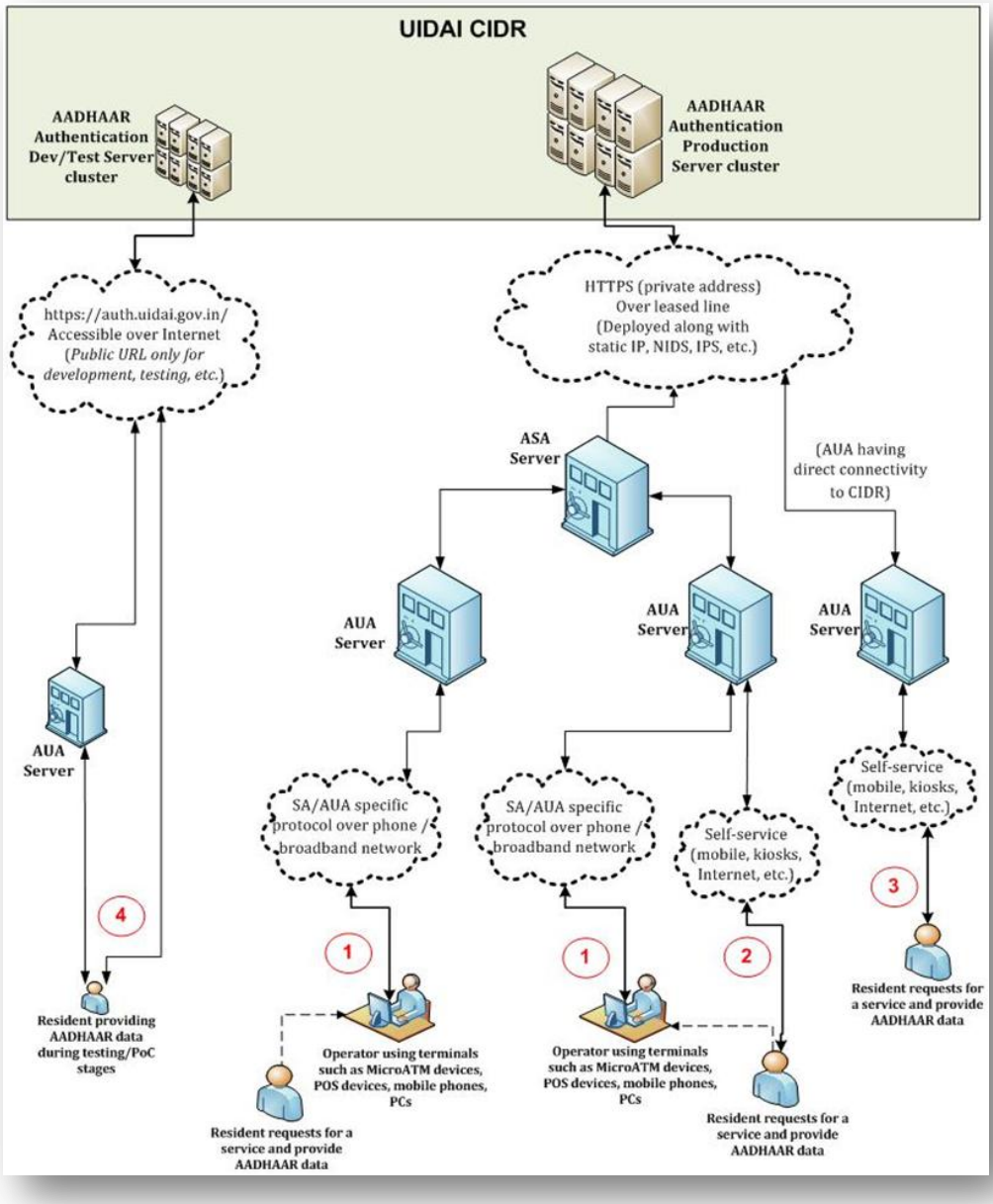


Figure 10 UIDAI Authentication Ecosystem

The diagram shows various configurations in which the ecosystem works together to authenticate a resident with the UIDAI.

11.5 Authentication Device Ecosystem

Authentication devices typically comprise a host, and a sensor. The host is available in many form factors; it could be a mobile phone, a point of sale device, a kiosk, an ATM, or

a PC. The host may support one or more applications, which allow users to get access to services. The biometric sensor is also available in many form factors, and could be an embedded module, or a separately packaged device.

To ensure standardization, and the highest quality authentication experience, with low error rates, the UIDAI engaged in extensive consultations, and POC studies. These resulted in the appointment of the STQC as the agency to certify biometric sensors that can be used with Aadhaar authentication.

11.6 Application Ecosystem

The UIDAI has encouraged the development of an application ecosystem, and provides a platform for the UIDAI authentication application developers to engage with each other, and with the UIDAI for technical support.

This is available as a part of the Authentication portal on the UIDAI website.

12 Biometrics in Authentication

The UIDAI uses biometrics as a 'Who You Are' factor for authentication. It is important to understand some aspects of biometrics as they pertain to this usage.

12.1 Biometrics for Authentication

Fingerprint and iris are the biometric modalities that are being used by UIDAI to allow residents to authenticate themselves. Online biometric authentication is a 1:1 verification of the biometric(s) presented at the time of authentication with templates generated from the data collected during enrolment or biometric updates.

Biometric matches are not exact matches, and such system may fail in the following two ways:

- **False Reject** - The authentication system incorrectly rejects a genuine claim of identity (authentication attempt). The False Reject Rate (FRR) is defined as ratio of the number of false rejects to the number of genuine authentication attempts.
- **False Accept** - The authentication system incorrectly accepts an imposter claim of identity. False Accept Rate (FAR) is defined as ratio of the number of false accepts to the number of impostor authentication attempts.

Similar to enrolment the accuracy of authentication may also be expressed using an ROC curve that shows the trade-off between FAR and FRR. One important difference between enrolment biometrics and authentication biometrics is while the resident presents all ten fingerprints and two iris images during enrolment only one or two fingerprint or iris is presented during authentication.

To attain the best results, the UIDAI makes the following recommendations to application developers:

- **Capture Quality**: Measure the quality of captured biometrics, and request the resident to re-present their biometrics for poor quality captures.
- **2-Finger Authentication**: Fusion allows 2-finger authentication to have lower false reject rate. Use it!
- **Multiple Attempts**: Allow the application to have multiple attempts, before rejecting the business transaction.
- **Backup Authentication**: Have an alternate form of authentication to allow users who may be unable to use biometrics for authentication.
- **Best Finger Detection**: Use Best Finger Detection to guide users to use the best fingers for authentication.

13 Electronic KYC

Verification of the Proof of Identity (PoI) and Proof of Address (PoA) is a key requirement for access to financial products (payment products, bank accounts, insurance products, market products, etc.), SIM cards for mobile telephony, and access to various Central, State, and Local Government services. Today, customers provide physical PoI and PoA documents. These requirements, from various domains, are collectively referred to as the 'Know Your Customer' (KYC) requirements. While these requirements are derived from various laws relating to the prevention of money laundering, financing of terrorist activities, etc. – they do become a point of friction for a customer approaching a service provider.

Aadhaar has already been notified as a valid PoI and PoA document for various services in the Financial, Telecom, and Government domains. Being an electronically available, online identity system, Aadhaar provides a convenient method to allow a service provider to meet its obligations, and rapidly on-board (even in real time) a customer for service.

13.1 E-KYC

Service providers can provide a paperless KYC experience and avoid the cost of repeated KYC, the cost of paper handling and storage, and the risk of forged documents. Authorized service providers may access the Aadhaar e-KYC service from UIDAI through the e-KYC API specified by the UIDAI [eKYC API].

The Aadhaar e-KYC API provides a convenient mechanism for agencies to offer an electronic, paper-less KYC experience to Aadhaar holders. The e-KYC service provides simplicity to the resident, while providing cost-savings from managing and processing paper documents to the service agencies. UIDAI has published its policy on e-KYC and API is implemented as per this.

13.2 Salient Features of E-KYC Service

1. **Paperless** - The service is fully electronic, and document management can be eliminated.
2. **Consent based** - The KYC data can only be provided upon authorization by the resident through Aadhaar authentication, thus protecting resident privacy.
3. **Eliminates document forgery** - Elimination of photocopies of various documents that are currently stored in premises of various stakeholders reduces the risk of identity fraud and protects resident identity. In addition, since the e-KYC data is provided directly by UIDAI, there is no risk of forged documents.
4. **Inclusive** - The fully paperless, electronic, low-cost aspects of e-KYC make it more inclusive, enabling financial inclusion.
5. **Secure and compliant with the IT Act** - Both end-points of the data transfer are secured through the use of encryption and digital signature as per the Information Technology Act, 2000 making e-KYC document legally equivalent to paper documents. In addition, the use of encryption and digital signature ensures that no unauthorized parties in the middle can tamper or steal the data.
6. **Non-repudiable** - The use of resident authentication for authorization, the affixing of a digital signature by the service provider originating the e-KYC request, and the affixing of a digital signature by UIDAI when providing the e-KYC data makes the entire transaction non-repudiable by all parties involved.
7. **Low cost** - Elimination of paper verification, movement, and storage reduces the cost of KYC to a fraction of what it is today.
8. **Instantaneous** - The service is fully automated, and KYC data is furnished in real-time, without any manual intervention.
9. **Machine readable** - Digitally signed electronic KYC data provided by UIDAI is machine readable, making it possible for the service provider to directly store it as the customer record in their database for purposes of service, audit, etc. without human intervention making the process low cost and error free.
10. **Regulation friendly** - The service providers can provide a portal to the Ministry/Regulator for auditing all e-KYC requests. The Ministry/Regulator can establish rules for secure retention of e-KYC data (e.g. storage method, period of storage, and retrieval among other things).

13.3 E-KYC Ecosystem

The Aadhaar e-KYC API can be used (only with the explicit authorization of the resident through biometric/OTP authentication) by an agency to obtain latest resident demographic data and photo data from UIDAI. The resident servicing agency is called the KYC User Agency (KUA). The KUA accesses the e-KYC service through a KYC Service Agency (KSA). The KSA provides connectivity to CIDR.

13.4 E-KYC Process

13.4.1 Use Cases

Broadly speaking, two scenarios under which the e-KYC service can be used:

1. New customer/beneficiary:
    a. The KUA captures resident authentication data and invokes the Aadhaar e-KYC API through a KSA network;
    b. The KYC data returned within the response of the e-KYC API is digitally signed and encrypted by UIDAI; and
    c. Using the resident data obtained through this KYC API, the agency can provision the service instantaneously.
2. Existing customer/beneficiary:
    a. The KUA captures resident authentication data and invokes the Aadhaar e-KYC API through a KSA network;
    b. The KYC data returned within the response of the e-KYC API is digitally signed and encrypted by UIDAI;
    c. Since the resident is already a customer/beneficiary, the KUA can use a simple workflow to approve the Aadhaar linkage by comparing data retrieved through the e-KYC API against what is on record (in paper or electronic form); and
    d. Once verified, the existing customer/beneficiary record can be linked to the Aadhaar number. The Aadhaar e-KYC API returns data along with a unique transaction code. The fact that the data is digitally signed by UIDAI and that every transaction has a unique code makes it possible to perform an electronic audit at a later point in time for any particular transaction.

13.4.2 Data Flow

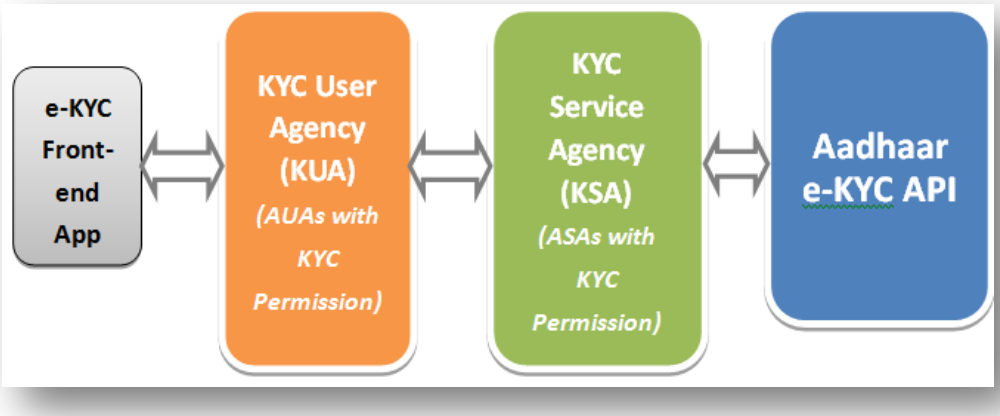


Figure 11: e-KYC Data Flow

Following the data flow of a typical KYC API call from left to right and back.

1. KYC front-end application captures Aadhaar number + biometric/OTP of resident and forms the encrypted PID block.
2. KUA forms the input XML, digitally signs it, and sends to KSA.
3. KSA forwards the KYC XML to Aadhaar KYC API.
4. After successful resident authentication, responds with digitally signed and encrypted XML containing demographic and photograph of the resident.
5. KSA sends the response back to KUA enabling paper-less electronic KYC.

13.5 Information Privacy & Security

The Aadhaar e-KYC service does not compromise security for inclusion, and instead offers a solution that is secure as well as inclusive and protects data privacy by eliminating paper trail on the field.

For further details, refer to "UIDAI policy on e-KYC service" [EKYC Policy, 2012] *and "Aadhaar e-KYC API 1.0"* [eKYC API] *documents.*

Part IV: Appendices

14 Language Support

The following table shows the default second language for each state / union territory at the time of enrolment.

State / Union Territory	*Language Supported*
Andaman and Nicobar Islands	Hindi
Andhra Pradesh	Telugu
Arunachal Pradesh	Assamese
Assam	Assamese
Bihar	Hindi
Chandigarh	Punjabi, Hindi, English
Chhattisgarh	Hindi
Dadra and Nagar Haveli	Marathi, Gujarati
Daman and Diu	Gujarati
Delhi	Hindi
Goa	Konkani (Marathi Script)
Gujarat	Gujarati
Haryana	Hindi
Himachal Pradesh	Hindi
Jammu and Kashmir	Urdu
Jharkhand	Hindi
Karnataka	Kannada
Kerala	Malayalam
Lakshadweep	Malayalam

State / Union Territory	*Language Supported*
Madhya Pradesh	Hindi
Maharashtra	Marathi
Manipur	Manipuri
Meghalaya	English
Mizoram	English
Nagaland	English
Orissa	Oriya
Pondicherry	Tamil for Puducherry, Malayalam for Mahe, Telugu for Yanam
Punjab	Punjabi
Rajasthan	Hindi
Sikkim	Nepali
Tamil Nadu	Tamil
Tripura	Bengali
Uttar Pradesh	Hindi
Uttarakhand	Hindi
West Bengal	Bengali

15 References

[Aadhaar Number, 2010]: A UID Numbering Scheme, 2010
http://uidai.gov.in/UID_PDF/Working_Papers/A_UID_Numbering_Scheme.pdf

[ABIS API, 2012]: Aadhaar Automated Biometric Identification Subsystem Interface
http://uidai.gov.in/UID_PDF/Working_Papers/Aadhaar_ABIS_API.pdf

[Analytics, 2012]: Analytics – Empowering Operations – The UIDAI Experience
http://uidai.gov.in/images/FrontPageUpdates/uid_doc_30012012.pdf

[Architecture, 2014]: Aadhaar Technology & Architecture: Principles, Design, Best Practices and Key Learnings
http://www.uidai.gov.in/images/AadhaarTechnologyArchitecture_March2014.pdf

[Authentication, 2012]: AADHAAR Authentication API Specification - Version 1.5
http://uidai.gov.in/images/FrontPageUpdates/aadhaar_authentication_api_1_5_rev1_1.pdf

[Authentication Accuracy, 2012]: Role of Biometric Technology in Aadhaar Authentication – Authentication Accuracy Report
http://uidai.gov.in/images/FrontPageUpdates/role_of_biometric_technology_in_aadhaar_jan21_2012.pdf

[Authentication Model, 2012]: AADHAAR Authentication Operating Model
http://www.uidai.gov.in/images/authDoc/d3_1_operating_model_v1.pdf

[BFD, 2012]: AADHAAR BEST FINGER DETECTION API Specification - Version 1.0
http://uidai.gov.in/images/FrontPageUpdates/aadhaar_bestfingerdetection_api_1.0_draft.pdf

[Biometric Committee, 2009]: Biometrics Design Standards for UID Applications, By UIDAI Committee on BIometrics
http://uidai.gov.in/UID_PDF/Committees/Biometrics_Standards_Committee_report.pdf

[Biometric SDK API, 2012]: Aadhaar Biometric SDK API Specification Version 2.0
http://uidai.gov.in/images/aadhaar_biometric_sdk_api_2_0.pdf

[Biometrics Report, 2012]: Role of Biometric Technology in Aadhaar Enrolment
http://uidai.gov.in/images/FrontPageUpdates/role_of_biometric_technology_in_aadhaar_jan21_2012.pdf

[DDSVP, 2009]: Demographics Data Standards and Verification Procedure Committee Report, By DDSVP Committee
http://uidai.gov.in/UID_PDF/Committees/UID_DDSVP_Committee_Report_v1.0.pdf

[Device API, 2010]: Aadhaar Biometric Capture Device API
http://uidai.gov.in/images/Aadhaar_Biometrics_Capture_Device_API_1_0.pdf

[eKYC API]: Aadhaar eKYC API Specification – Version 1.0 (Final)
http://uidai.gov.in/images/aadhaar_kyc_api_1_0_final.pdf

[EKYC Policy, 2012]: UIDAI E KYC Server Policy
http://uidai.gov.in/images/ekyc_policy_note_18122012.pdf, 2012

[Enrolment PoC, 2010]: UID Enrolment Proof Of Concept Report
http://uidai.gov.in/images/FrontPageUpdates/uid_enrolment_poc_report.pdf

[IRIS Accuracy, 2012]: Role of Biometric Technology in Aadhaar Authentication –IrisAuthentication Accuracy Report
http://uidai.gov.in/images/detailed_poc_10_report_ver12a_23052013.pdf

[IRIS Accuracy, 2013]: Role of Biometric Technology in Aadhaar Authentication – Kind 7 IrisAuthentication Accuracy – PoC Report
http://uidai.gov.in/images/detailed_poc_10_report_ver12a_23052013.pdf

[Iris Inclusion, 2010]: Ensuring Uniqueness: Collecting iris biometrics for the Unique ID Mission
http://uidai.gov.in/UID_PDF/Working_Papers/UID_and_iris_paper_final.pdf

[ISO 19795-2, 2007]: Biometric performance testing and reporting—Testing methodologies for technology and scenario evaluation. By International Standards Organization.

[Jain, 1999]: Jain, Prabhakar and Ross, 1999
http://www.csee.wvu.edu/~ross/pubs/RossFingMatch_MSUTR99-14.pdf

[NIST, 2006]: MINEX Performance and Interoperability of the INCITS 378 Fingerprint Template, Supplement No. 1 Native Matching, Patrick Grother, Michael McCabe, Craig Watson, Mike Indovina, Wayne Salamon, Patricia Flanagan, Elham Tabassi, Elaine Newton, Charles Wilson, National Institute of Standards and Technology March 21, 2006

[NISTIR 7346]: NISTIR 7346, Studies of Biometric Fusion, Brad Ulery, Austin Hicklin, Mitretek Systems, Craig Watson Image Group, Information Access Division Information Technology Laboratory, William Fellner, Mitretek Systems, Peter Hallinan, Mitretek Systems Consultant

[Notification, 2009]: Notification constituting the Unique Identification Authority of India

[Security]: Aadhaar Security Policy & Framework for Aadhaar Authentication
http://uidai.gov.in/images/authDoc/d3_4_security_policy_framework_v1.pdf

[Service Delivery]: Aadhaar Enabled Service Delivery
http://uidai.gov.in/images/authDoc/whitepaper_aadhaarenabledservice_delivery.pdf

[STQC, Authentication, 2011]: STQC: UIDAI Biometric Authentication Device specification
http://stqc.gov.in/sites/upload_files/stqc/files/STQC UIDAI BDCS-03-08 UIDAI Biometric Device Specifications_Authentication_1.pdf

[STQC, Certification, 2011]: STQC Biometric Devices Testing and Certification
http://www.stqc.gov.in/content/bio-metric-devices-testing-and-certification

[Strategy, 2009]: Creating a Unique Identity for All Residents - A Draft Approach
http://www.uidai.gov.in/documents/Creating_a_unique_identity_for_every_resident_in_India.pdf

[Strategy, 2014]: Aadhaar Technology Strategy, Ecosystem, Technology and Governance
http://www.uidai.gov.in/images/AadhaarTechnologyStrategy_March2014.pdf

[Update Policy, 2013]: UIDAI Data Update Policy, Ver 2.3
http://uidai.gov.in/images/mou/uidai_data_update_policy_ver_2.3.pdf

[Wayman, 2002]: Best practices in testing and reporting performance of biometric devices, Version 2.01 By A. J. Mansfield, National Physical Laboratory and J. L. Wayman, San Jose State University. Middlesex: NPL Report CMSC 14/02.

Published by UIDAI Technology Center
Bengaluru, Karnataka, India.

(c) UIDAI, 2014, All Rights Reserved.