

**IDENTITY SCHEMES FOR CITIZENS:  
LESSONS FROM INDIAN EXPERIENCE OF  
'AADHAAR'**

**RAM SEWAK SHARMA**



**AMAR NATH & SHASHI KHOSLA SCHOOL OF INFORMATION  
TECHNOLOGY**

**INDIAN INSTITUTE OF TECHNOLOGY DELHI  
FEBRUARY 2018**

© **Indian Institute of Technology, Delhi (IITD), New Delhi, 2018**

# **IDENTITY SCHEMES FOR CITIZENS: LESSONS FROM INDIAN EXPERIENCE OF 'AADHAAR'**

by

**RAM SEWAK SHARMA**

Amar Nath & Shashi Khosla School of Information Technology

Submitted

in fulfilment of the requirements of the degree of Doctor of Philosophy

to the



**INDIAN INSTITUTE OF TECHNOLOGY DELHI**

FEBRUARY 2018

## Certificate

This is to certify that the thesis entitled “**Identity Schemes for Citizens: Lessons from Indian experience of ‘Aadhaar’**” being submitted by **Ram Sewak Sharma** to the Indian Institute of Technology Delhi for the award of the degree of **Doctor of Philosophy (PhD)** is a record of *bona fide* research work carried out by him. He has worked under my guidance and supervisions and fulfilled the requirements for the submission of the thesis, which has attained the standard required for a PhD degree of the Institute. The results presented in this thesis have not been submitted elsewhere for the award of any degree or diploma.

**(Prof. MP Gupta)**

**Research Supervisor**

Department of Management Studies,  
Indian Institute of Technology Delhi

## **Acknowledgements**

First and foremost, I would like to express my sincere thanks and appreciation to my supervisor Prof (Dr) MP Gupta for his patient and professional guidance towards shaping this PhD thesis from disjointed facts and material to an organically woven research work.

I would also like to acknowledge the support of a large number of my acquaintances and other participants who voluntarily devoted their time to answer the questions on various aspects of UID Project and provided their valuable comments and insights. While I am not writing their names here (their number being around 100), it in no way reduces their contribution to this research and my gratitude to them. I am also grateful to many of my UIDAI colleagues for sharing valuable information about the project.

I also want to express my sincere gratitude to those of my UIDAI colleagues and others who participated in the expert group discussion about the goals and design/execution choice discussion. They are: Mrs. Sujata Chaturvedi, Mrs. Keshni Anand Arora, Mr. Anil Khachi, Ms. Deepti Vikas Dutt and Mr. SK Barnawal.

I also want to acknowledge the help rendered by some of my colleagues (Mr. Anil Vishnoi, Ms. Sakshi Singh, Ms. Shiuli Bhattacharya, Ms. Shweta Singh, Mr. Rajesh Sharma and Mr. Prakash Kirtani) in my office. Though this was not part of their official duties, they willingly came forward and helped me in activities like formatting, proof-reading and pagination of this thesis.

Finally, I would like to acknowledge the support of my wife Kalpana for providing me her silent support and encouragement.

**(RS Sharma)**

## **Abstract**

This research aims to study various challenges around identify scheme ‘Aadhaar’ that have been raised since its inception in 2009. Aadhaar was started by the Government of India (GoI) as Unique Identification project. Two major reasons guided this project. Lack of an identity document with a large number of people, resulting in their exclusion from accessing government services and benefits was the first. Existence of duplicates and ghosts in many databases resulting in huge leakages in various benefit delivery programs was the second. The first reason mandated the system to be inclusive and the second one required that the identities to be unique to eliminate duplicate and ghosts.

A project for creating unique identities for 1.2 billion people was both unique and challenging. Uniqueness was to be achieved using biometric technology. Nowhere in the world has a project of this size has ever been attempted. In addition to ensure uniqueness, the project promised to provide authentication service, a facility where you could prove who you are, anytime and anywhere, using your biometric credentials. Most importantly, the project was to be built as a platform, so as to be used for multiple domains. It promised to solve the common problem of establishing identity before any service delivery, something almost every service delivery domain required.

Thus, Aadhaar is a massive technology project. It is also a project having significant public policy and governance implications. Various aspects, including and not limited to, its relevance and context, its design and implementation, technology architecture, logistics, coordination with other agencies, criticisms and the responses, application eco-system, privacy issues and its service delivery potential become very interesting candidates for study. Researching these issues not only provides an insight into implementation complexities of this project, it also provides important lessons for managing massive ICT projects.

This study started with examination of the concept of identity and how this can be used in the service delivery. It has examined the over-arching goals of Aadhaar and the design choices which the project made for achieving those goals. As technology is central to Aadhaar implementation, it has examined the technical architecture of this project. It has also gone into the application eco-system and has suggested algorithms to expedite the seeding of ID numbers into various databases.

It is pertinent to examine feasibility of using biometrics at such a scale and the appropriateness of the architecture to deploy these solutions. It has also carried out research around design of the project and issues like cost of technology and vendor lock-in issues.

Aadhaar has also attracted serious criticisms from privacy angle. This has been examined in detail, particularly the claim of UIDAI about embedding the privacy in the design of UID project.

Aadhaar is a very complex and sophisticated project having interplay of technology, public policy, inter-agency coordination and logistics at unprecedented scales. When this project started getting implemented, many challenges arose which were neither thought of nor could have been conceived. Creating and rolling out solutions to these problems required ingenuity, understanding of technology by the project team and speed. Attempt has been made to examine the maturity levels of various sub-systems around implementation of this project and other challenges which arise out of policy and strategic decisions of UIDAI, the authority responsible for implementing this project.

The thesis also presents a detailed analysis of innovations which have been introduced in this project. Being a technology project being implemented in a diverse country like India, this could not have been implemented without mid-course innovations to deal with unforeseen challenges.

Major research findings include the challenges around this project and how it has dealt with them, whether the approach was the most appropriate one or not. One major research contribution is development of a methodology wherein each project, especially an ICT project should have clearly articulated goals and then should examine every design and implementation choice against the touch-stone of the over-all project goals. The methodology developed in this research relating to goals and design choices could be further fine-tuned so as to be used with much more precise and quantitative results. Another major contribution is the seeding algorithm developed which could accelerate the deployment of Aadhaar in various applications. Innovations in ICT projects having large logistics is another useful study. Study of technology choices in Aadhaar provides valuable lessons as to how to design open and standard-based solutions so as to be scalable and cost-effective.

While this research is focused on one project it has important implications and lessons for practitioners and project managers. Some of these are: (a) Appropriate agencies should come

out with code of best practices for eGovernance projects., (b) All eGovernance project attempt to simplify processes and make it easy for citizens to access government services. Body of knowledge gained through study of various eGovernance projects must be used to fine-tune the design principles of such projects. (c) There is a need to develop some algorithms and methods to tackle the issue of immutability of biometrics which is one of the major concerns in this project. (d) More research needs to be carried out in the area of phonetic algorithms for Indo-Aryan languages. More active role should be played in the area of biometrics research and standard setting to strengthen such project as it has used biometrics for development and inclusion for the first time at such a scale.

## सार

इस अध्ययन का उद्देश्य पहचान योजना 'आधार' की विभिन्न चुनौतियों का अध्ययन करना है जिन्हें वर्ष 2009 में इसके आरंभ से ही उठाया जा रहा था। आधार को भारत सरकार (जीओआई) द्वारा विशिष्ट पहचान परियोजना के रूप में आरंभ किया गया था। इस परियोजना को दो महत्वपूर्ण कारणों के चलते आरंभ किया गया था। सर्वप्रथम, बड़ी संख्या में लोगों के पास पहचान के दस्तावेज का न होना था जिसके कारण उन्हें सरकारी सेवाओं तथा लाभों से वंचित होना पड़ा था। दूसरे, अनेक डाटाबेस में लोगों का एक बार से ज्यादा मौजूद होना तथा अस्तित्वहीन व्यक्तियों द्वारा विभिन्न लाभ प्रदान करना था जिसके परिणामस्वरूप कार्यक्रमों में बड़ी हेराफेरी का होना था। प्रथम कारण के परिणामस्वरूप प्रणाली को समेकित किया जाना अधिदेशित था तथा दूसरे कारण के परिणामस्वरूप एक बार से ज्यादा मौजूद तथा अस्तित्वहीन व्यक्तियों को हटाने के लिए पहचान विशिष्ट होनी अधिदेशित थी।

1.2 बिलियन लोगों के लिए विशिष्ट पहचान तैयार करना, विशिष्ट तथा चुनौतिपूर्ण था। बायोमैट्रिक प्रौद्योगिकी का उपयोग करके इस विशिष्टता को प्राप्त किया जाना था। विश्व में कहीं भी इस आकार की परियोजना को लागू करने का कभी प्रयास भी नहीं किया गया। विशिष्टता सुनिश्चित करने के अतिरिक्त, परियोजना के माध्यम से अधिप्रमाणन सेवा उपलब्ध कराई जानी थी, जोकि एक सुविधा है जहां आप कहीं भी, कभी भी अपने बायोमैट्रिक प्रत्यायकों का उपयोग कर सिद्ध कर सकते हैं कि आप कौन हो। सबसे महत्वपूर्ण बात यह है कि परियोजना को एक प्लेटफार्म के रूप में तैयार किया जाना था ताकि इसे बहु प्रणालियों के लिए उपयोग किया जा सके। इसमें किसी सेवा सुपुदुर्गी के लिए पहचान सिद्ध करने की सामान्य समस्या का समाधान किया जाना था, जोकि लगभग सभी प्रणालियों में सेवा सुपुदुर्गी के लिए अपेक्षित था।

अतः आधार एक बृहद प्रौद्योगिकी परियोजना है। यह एक ऐसी भी परियोजना है जिसके पर्याप्त सार्वजनिक नीतिगत तथा अभिशासनात्मक परिणाम हो सकते हैं। असीमित रूप से इसकी संगतता तथा संदर्भ इसका डिजाइन, कार्यान्वयन, प्रौद्योगिकीय ढांचा, लाजिस्टिक्स, अन्य एजेन्सियों के साथ समन्वय तथा प्रतिक्रिया, अनुप्रयोग का परिवेश, निजता के मुद्दे तथा सेवा सुपुदुर्गी क्षमता संबंधी विभिन्न पहलू अध्ययन हेतु अभ्यर्थियों के लिए काफी रूचिकर हो सकते हैं। इन मुद्दों के संबंध में शोध करने से न केवल परियोजना की कार्यान्वयन संबंधी जटिलताओं का पता चलता है बल्कि यह वृहद आईसीटी परियोजनाओं का प्रबंधन करने के लिए महत्वपूर्ण सीख देता है।

यह अध्ययन पहचान की अवधारणा की जांच तथा इसे कैसे सेवा सुपुदुर्गी में उपयोग किया जा सकता है, से आरंभ हुई। इसमें आधार के अति महत्वपूर्ण उद्देश्यों तथा डिजाइन

संबंधी विकल्पों, जिसे इन उद्देश्यों की प्राप्ति हेतु परियोजना में सम्मिलित किया गया, की जांच की है। चूंकि प्रौद्योगिकी, आधार कार्यान्वयन का केन्द्र बिंदु है। इसमें इस परियोजना के तकनीकी ढांचे की भी जांच की गई है। इसमें अनुप्रयोग संबंधी परिवेश के माध्यम से विभिन्न डाटाबेस में आईडी नम्बरों को जोड़ने हेतु एलाग्रिथम का भी सुझाव दिया गया है।

इतने बड़े पैमाने पर बायोमैट्रिक के उपयोग की व्यवहार्यता की जांच करने तथा इन समाधानों का उपयोग करने के लिए ढांचे की उपर्युक्तता की जांच करना महत्वपूर्ण है। इसमें परियोजना के डिजाइन के माध्यम से प्रौद्योगिकी की लागत तथा वेंडर-लॉक-इन जैसे मुद्दों पर शोध भी की गई है। आधार की निजता के दृष्टिकोण से भी गंभीर आलोचनाएं की गई हैं। इनकी विस्तार से जांच की गई है, विशेष रूप से यूआईडी की परियोजना के ढांचे में निजता को शामिल करने के यूआईडीएआई के दावे की जांच की गई है।

आधार एक अत्यंत जटिल तथा अत्याधुनिक परियोजना है जिसमें अनेक प्रौद्योगिकीयों, जन नीति, अंतर-एजेन्सी समन्वय तथा लाजिस्टिक्स का अभूतपूर्व एवं व्यापक स्तर पर परस्पर मेल-जोल से कार्य करना शामिल है। जब इस परियोजना को लागू किया जाना आरंभ किया, अनेक चुनौतियां उभरीं जिनके बारे में न तो सोचा गया था न ही परिकल्पना की गई थी। इन समस्याओं के लिए समाधान तैयार करने तथा उसे लागू करने के लिए कल्पनाशीलता, परियोजना दल द्वारा परियोजना का बोध तथा तीव्रता से कार्य किए जाने की आवश्यकता थी। इन परियोजनाओं के कार्यान्वयन के साथ विभिन्न उप-प्रणालियों की परिपक्वता के स्तर तथा अन्य चुनौतियों जैसे यूआईडीएआई, जोकि इस परियोजना को लागू करने के लिए उत्तरदायी एजेन्सी थी, की नीतिगत तथा रणनीतिक निर्णयों के परिणाम-स्वरूप उभरी, की जांच करने के लिए प्रयास किए गए।

यह प्रबंध शोध ऐसे नवोन्मेषों का विश्लेषण भी प्रस्तुत करता है जिन्हें इस परियोजना में शामिल किया गया है। भारत जैसे देश में कार्यान्वित की गई प्रौद्योगिकीय परियोजना होने के नाते, इसे अप्रत्याशित चुनौतियों का सामना करने के लिए मध्यावधि नवोन्मेषों के बिना कार्यान्वित नहीं किया जा सकता था।

वृहद शोध निष्कर्षों में इस परियोजना से जुड़ी चुनौतियां तथा उनसे कैसे निपटा गया, क्या यह पद्धति उपर्युक्त थी अथवा नहीं, शामिल हैं। वृहद शोध में, एक पद्धति को विकसित किया जाना, एक योगदान था जहां प्रत्येक परियोजना, विशेष रूप से आईसीटी परियोजना के स्पष्ट रूप से चिह्नित उद्देश्य होने चाहिए तत्पश्चात् समग्र परियोजना उद्देश्यों के मापदंड के समक्ष प्रत्येक डिजाइन तथा कार्यान्वयन संबंधी विकल्पों की जांच की जानी चाहिए थी। इस शोध में उद्देश्यों तथा डिजाइन संबंधी विकल्पों से जुड़ी पद्धतियों को आगे और अधिक परिशोधित

किया जा सकता है जिससे इन्हे अधिक सटीकता तथा परिमाणात्मक परिणामों के साथ उपयोग किया जा सकता है। अन्य महत्वपूर्ण योगदान में, विकसित की गई एलाग्रिथम को सीडिंग करना था जिससे विभिन्न अनुप्रयोगों में आधार के उपयोग को गति दी जा सके। वृहदाकार लाजिस्टिक्स वाली आईसीटी परियोजनाएं एक अन्य लाभप्रद अध्ययन है। आधार में प्रौद्योगिकीय विकल्पों के अध्ययन से महत्वपूर्ण निष्कर्ष निकलते हैं जैसेकि मानक आधारित समाधानों को किस प्रकार डिजाइन किया जाए, खोला जाए ताकि उनके दायरे को व्यापक बनाया जा सके तथा उन्हें लागत प्रभावी बनाया जा सके।

जबकि यह शोध एक परियोजना पर आधारित है तथापि, इसके इस क्षेत्र के पेशेवरों तथा परियोजना प्रबंधकों के लिए इसके महत्वपूर्ण निष्कर्ष तथा परिणाम निकले हैं। इनमें से कुछ नीचे दिए गए हैं: (क) उपर्युक्त एजेन्सियों को ई-गवर्नेंस परियोजनाओं के लिए सर्वश्रेष्ठ पद्धतियां विकसित करनी चाहिए, (ख) सभी ई-गवर्नेंस परियोजनाओं का प्रयास प्रक्रियाओं का सरलीकरण किया जाना तथा नागरिकों को सरकारी सेवाओं तक पहुंच बनाने को सुगम बनाना है। विभिन्न ई-गवर्नेंस परियोजनाओं के अध्ययन के माध्यम से प्राप्त ज्ञान को ऐसी परियोजनाओं के डिजाइन सिद्धांतों को ठीक करने के लिए इस्तेमाल किया जाना चाहिए। (ग) बायोमेट्रिक्स की अपरिवर्तनीयता के मुद्दे के समाधान के लिए कुछ एल्गोरिदम और विधियों को विकसित करने की आवश्यकता है जो कि इस परियोजना में मुख्य चिंताओं में से एक है। (घ) इंडो-आर्यन भाषाओं के लिए ध्वन्यात्मक एल्गोरिदम के क्षेत्र में अधिक शोध किए जाने की आवश्यकता है। बायोमेट्रिक्स अनुसंधान और मानक सेटिंग के क्षेत्र में और अधिक सक्रिय भूमिका निभाई जानी चाहिए ताकि ऐसी परियोजना को मजबूत किया जा सके, क्योंकि इसने इस तरह के पैमाने पर पहली बार विकास और समावेशन के लिए बायोमेट्रिक्स का इस्तेमाल किया है।

# TABLE OF CONTENTS

Content	Page No.
Certificate	i
Acknowledgement	ii
Abstract	iii
List of Figures	iii
List of Tables	xvii
List of Appendices	xxi
List of Abbreviations Used	xxii
Chapter 1: Introduction to the Study	1
1.1 Background	1
1.2 Events leading to the constitution of UIDAI	2
1.3 Need for one more ID document	4
1.4 Motivation for research	5
1.5 Outline of research	7
1.6 Organization of the Report	8
Chapter 2: Literature Review	12
2.1 Introduction:	12
2.2 What is Identity?	12
2.2.1 Social and Political Science perspectives:	12
2.2.1 Security perspective:	13
2.2.3 National ID Cards and similar documents	13
2.2.4 Identity attributes and eligibilities	14
2.2.5 Digital Identities	15
2.3 Role of Identity in governance	17
2.4 Identity: The Indian Context	18
2.5 Challenges in the existing identity documents	20
2.6 Need for Unique IDs in India	21

2.7	Unique ID Project of India	22
2.8	Research Gaps	22
2.9	Concluding Remarks	24
Chapter 3: Design of the Study		25
3.1	Introduction	25
3.2	Problem Statement	25
3.3	Research Questions	26
3.4	Research Objectives	28
3.5	Research Methodology	29
3.6	Concluding remarks	33
Chapter 4:Analysing Design and Execution Choices made to support overarching goals of Aadhaar		35
4.1	Introduction	35
4.2	Approach and methodology of the study	35
4.3	Expert Group Methodology	36
4.4	Members of the Expert Group	36
4.5	Expert Group Deliberations	38
4.6	Design and Execution Choices	38
4.7	Goals of UID Project and the support levels of Design and Execution Choices	38
4.7.1	Uniqueness of Identities	39
4.7.2	Inclusion	42
4.7.3	Resident Convenience	45
4.7.4	Cost Optimisation	47
4.7.5	Speed	49
4.7.6	Scalability	51
4.7.7	Building it as a platform	53
4.7.8	Future-proofing the Technology	54
4.7.9	Technology: Data Security and Privacy	55
4.8	Consistency of the set of over-arching goals	56
4.9	Over-all Scores	57

4.10	Concluding Remarks	58
Chapter 5: Aadhaar Technology (Biometrics)		60
5.1	Introduction	60
5.2	Problem Statement	60
5.3	Objective of the Study	61
5.4	Setting the framework	61
5.5	Iris: the critical biometric	63
5.6	Inclusion of Iris	65
5.7	Criticism relating to biometrics	66
5.7.1	Inaccuracy in de-duplication	66
5.7.2	Scalability of de-duplication	68
5.7.3	Inaccuracy and feasibility of Authentication	69
5.8	Reducing the risks	70
5.9	Accuracy of de-duplication	72
5.10	Feasibility of Authentication	73
5.11	Scalability of Authentication and other related operations	75
5.12	Concluding Remarks	80
Chapter 6: Aadhaar seeding through dictionary of tokens		82
6.1	Introduction	82
6.2	Context Description	82
6.3	Problem Statement	84
6.4	Name matching Algorithms and Approaches	86
6.4.1	Phonetic Matching Algorithms:	87
	Soundex:	88
	Metaphone:	89
	Beider-Morse Phonetic Matching	90
	Phonex:	90
6.4.2	String Matching Algorithms:	91
	Guth Name Matching (Guth, 1976)	91

k-Approximate String Matching	93
6.5    Algorithms used in the implementation their rationale	93
Step I: Creation of a Token Dictionary:	94
Step II: Creation of an Excluded Tokens' List	96
Step III: Removing very low frequency tokens from tokens list:	96
Customizing String/Token Matching Algorithms	97
Need to use second-stage filtering of tokens	98
6.6    Working of the Algorithm	99
Logic of the Algorithm:	101
Results of running the algorithms with different parameters	102
Method of Comparison of results:	103
6.7    Analysis of the Results and Conclusions	104
Results of exact match:	109
6.8    Matching Simulator	111
6.9    Optimizations which will improve performance	113
6.9.1    Creating a dictionary/list of 'fused' tokens	113
6.9.2    Creating a priority processing list	114
6.9.3    Differentiating Names and surnames	114
6.9.4    Learning from the data	115
6.9.5    Using more appropriate data structures	115
6.10   Concluding Remarks	115
Chapter 7: Aadhaar Maturity Analysis	117
7.1    Introduction	117
7.2    Methodology	117
7.3    The Experts	118
7.4    Issues around Aadhaar	122
7.5    Scope of the study	122
7.5.1    Enrolment Related	122
i.    Data capture errors during enrolment – Technology/Process	123

ii.	Integrity of Aadhaar enrolments and credibility of data collection methods – Technology/Process	126
iii.	Inclusion of people with biometric exceptions – Technology/Process	130
iv.	Updation of dynamic data – Technology/Process	132
v.	Standardization of the data collection process – Technology/Process	133
vi.	Biometrics: Collection, quality and scale issues – Technology/Process	135
vii.	Potential stealing and leakage of data by enrolment agencies or their agents. Mischief by EAs relating to multiple enrolments of the same person: Technology/Security	137
viii.	Enrolments: challenges of reaching remote areas, selective enrolments, residual population and leaving out the needy – Technology/Process	140
ix.	Delay in getting Aadhaar letter	142
	Fig. 7.2: Maturity graph of issues relating to enrolment eco-system.	145
7.5.2	Security and Privacy related issues	146
i.	Safety and Security of Biometric Database	146
ii.	Identity theft: immutability of biometrics	147
iii.	Potential misuse of data by Enrolment Agencies	149
iv.	Security vulnerability due to PKI ownership issues	150
7.5.3	Robustness of Technology	152
i.	Uniqueness of IDs	152
ii.	Authentication failure - denial of service:	154
iii.	Robustness of biometric authentication: who bears the consequences of false acceptance or false rejections	155
iv.	Security and Robustness Authentication on Internet – How safe and Secure?	156
7.6	Other issues and challenges around UID Project	160
7.6.1	Concerns relating to centralisation of data:	160
7.6.2	Threat of sharing of data with the Crime Investigating Agencies	162
7.6.3	Aadhaar violates privacy	163
7.6.4	No facility to de-enrol and delete the data:	166
7.6.5	Private Sector driven service delivery:	168
7.6.6	Should UIDAI not issue a smart card?	170

7.6.7	Aadhaar becomes the means of exclusion:	173
7.7	Policy related Issues	176
7.7.1	Aadhaar compromises national security by legitimising illegal immigrants	176
7.7.2	Aadhaar does not have a legal foundation	178
7.7.3	Duplication of work between NPR and UIDAI and wastage of public funds:	180
7.7.4	Aadhaar: Optional or Compulsory?	182
7.8	Concluding Remarks	185
Chapter 8: Aadhaar and Privacy		187
8.1	Introduction	187
8.2	Scope of the Study	188
8.3	Existing legal framework on data protection and privacy in India	189
8.3.1	Approach articulated in UIDAI's strategy document	190
8.4	Data Security and Privacy –in a digital world	191
8.5	Guiding approach to data protection and privacy	193
8.6	Considerations of Privacy and Data Security in Design, Implementation and legal framework of Aadhaar	197
I.	Minimal data collection	198
II.	Data use limitation	199
III.	Keeping the resident informed about her rights and data-usage	200
IV.	Resident consent and information for authentication	201
V.	Random Numbers with no intelligence	202
VI.	Stringent Data Sharing Policies and legal restrictions	203
VII.	Protection of Information and Accountability Mechanisms	205
VIII.	Exceptions to the data sharing provisions	205
IX.	Federated data model	206
X.	Data Protection Technologies: End-to-end Encryption	207
XI.	Putting the user in control	207
XII.	Access to the resident her own data	208
XIII.	Prohibition on publication	208

XIV.	Additional Measures relating to data and privacy protection	209
8.7	Analyzing Compliances of ‘Aadhaar’ over Privacy by Design (PbD) principles	209
i.	Proactive not Reactive; Preventative not Remedial	210
ii.	Privacy as the Default Setting	210
iii.	Privacy Embedded into Design	211
iv.	Full Functionality – Positive-Sum, not Zero-Sum	211
vi.	Visibility and Transparency – Keep it Open	213
vii.	Respect for user-privacy – Keep it user-centric	214
8.8	Revisiting Criticisms of the UID Project on Privacy	214
8.8.1	Major categories of criticisms and concerns privacy and measures adopted by UIDAI	215
8.8.2	An analysis of recent incidents concerning data security reported in press	216
i)	Massive “data-leak” of Aadhaar Numbers	216
ii)	Disclosure of Enrolment Information of MS Dhoni	218
iii)	Illegal use of Biometrics	219
8.9	Concluding Remarks	220
	Chapter 9: Analysis of Innovations in UID Project	222
9.1	Introduction	222
9.2	Scope of the Study	222
9.3	Innovations in the pre-enrolment eco-system:	222
9.3.1	Dual Screen in the enrolment machines	223
9.3.2	Only Certified Operators permitted to enrol residents	224
9.3.3	Standardization and validation of address entries:	225
9.3.4	Smart Enrolment Screens:	227
9.3.5	Quality check of biometrics at capture time:	228
9.3.6	Collection of Meta-data relating to enrolment:	228
9.3.7	Correction Window of 48 hours:	228
9.3.8	Outcome-based payment system	229
9.4	Post-enrolment Innovations	229
9.4.1	Online uploading of packets to CIDR	230

Aadhaar SFTP/Upload Portal	230
Aadhaar SFTP Client	232
Upload Portal (upload.uidai.gov.in)	232
9.5 Quality Control at the backend (QA Tool):	233
Checking of potential duplicates	236
9.6 Multiple Biometric Service Providers	238
9.7 Fake Aadhaar numbers – via biometric exception route	239
9.8 Innovation at Post-Aadhaar generation Stage	240
9.9 Last Mile Logistics (Aadhaar Print Factory)	240
9.10 Remote Aadhaar Seeding Framework	243
The problem	243
The solution: RASF	243
9.11 e-Aadhaar	246
9.12 Other innovations to check duplicate Aadhaar issuance	247
Concluding Remarks	248
Chapter 10: Conclusion	249
10.1 Introduction	249
10.2 Summary of the Study	249
10.3 Revisiting Research Questions	252
10.4 Revisiting Research Objectives (Presented below in Table 10.1)	254
10.5 Major research findings	255
10.6 Implications for policy makers and practitioners	258
10.7 Implication for researchers	259
10.8 Major research contributions	259
10.9 Limitations of our research	260
10.10 Future Scope of research	260
10.11 Concluding remarks	261
References	262
Appendix ‘A’	270
Appendix ‘B’	276
Appendix ‘C’	278

## List of Figures

<b>Figure Number</b>	<b>Description</b>	<b>Page Number</b>
Figure 1.1	Organization of thesis chapters	11
Figure 3.1	Flow of the thesis chapters	34
Figure 4.1	Impact of Design Choices on Uniqueness	40
Figure 4.2	Impact of Execution Choices on Uniqueness	40
Figure 4.3	Impact of Design Choices on Inclusion	44
Figure 4.4	Impact of Execution Choices on Inclusion	44
Figure 4.5	Design Choices having an impact on Resident Convenience	45
Figure 4.6	Execution Choices having an impact on Resident Convenience	46
Figure 4.7	Design Choices having an impact on Cost optimization	47
Figure 4.8	Execution Choices having an impact on Cost optimization	48
Figure 4.9	Design Choices having an impact on Speed	49
Figure 4.10	Execution Choices having an impact on Speed	50
Figure 4.11	Execution Choices having an impact on Scalability	52
Figure 4.12	Design Choices having an impact on Platform Approach	53
Figure 4.13	Design Choices having an impact on Future-proofing	54
Figure 4.14	Design Choices having an impact on Safety, Security and privacy	56

Figure 4.15	Execution Choices having an impact on Safety, Security and privacy	56
Figure 4.16	Over-all scores of Support Levels	58
Figure 5.1	Authentication transactions handled by UIDAI (month-wise)	78
Figure 5.2	eKYC transactions in the last four years.	79
Figure 5.3	eAadhaar downloads	79
Figure 6.1	Beneficiary data-table interacts with SRDH through Tokens data table	95
Figure 6.2	Distribution of tokens in various phonetic algorithms	98
Figure 6.3	How number of tokens grow with the number of SRDH records processed	101
Figure 6.4	Comparative results of various algorithms: Scholarship Bokaro Distt.	106
Figure 6.5	Comparative results of Success Ratio: Scholarship Bokaro Distt.	106
Figure 6.6	Comparative results of False Positive Ratio: Scholarship Bokaro Distt.	107
Figure 6.7	Comparative results of False Negative Ratio: Scholarship Bokaro Distt.	107
Figure 6.8	Comparative results of Various Algorithms: NSAP Bokaro Distt.	110
Figure 6.9	Comparative results of Success Ratio: NSAP Bokaro Distt.	111
Figure 6.10	Comparative results of False Positive Ratio: NSAP Bokaro Distt.	111
Figure-6.11	Screenshot of simulator providing the working of algorithm	112
Figure-6.12	Screenshot of simulator providing the parametrized view	113
Figure-7.1	Maturity stages	121
Figure-7.2	Enrolment Process and Technology	145
Figure-7.3	Data Security and privacy Issues	151
Figure-7.4	Robustness of Technology	159
Figure 8.1	Letter of Nandan Nilekani	192
Figure 8.2	Seven foundational principles of Privacy by Design	196
Figure- 9.1	Screenshot of data quality check screen for the QA Operator	235

Figure-9.2	Screenshot of data quality check screen for the QA Operator	235
Figure-9.3	Screenshot of data quality check screen for the QA operator	236
Figure-9.4	Screen-shot of potential duplicates in QA System of UIDAI	237
Figure-9.5	Screen-shot of QA MIS	238
Figure-9.6	Screen-shot of Seeding verification screen	245
Figure-9.7	Screen-shot of RASF verification process	246

## List of Tables

<b>Table Number</b>	<b>Description</b>	<b>Page Number</b>
Table 2.1	Estimated number of Cards and Card holders in India	19
Table 3.1	Research design	27-28
Table 6.1	Comparison pattern in Guth Name Matching	91
Table 6.2	Frequency distribution of tokens	96
Table 6.3	Partitions and Tokens created in various phonetic algorithms	97
Table 6.4	Results of Various Algorithms with different parameters Scholarship Data- Bokaro	105
Table 6.5	Results of Various Algorithms with different parameters: NSAP Bokaro	109-110
Table 7.1	Innovations in the enrolment system	125
Table 7.2	Responses in relation to the innovations employed by the UIDAI in the enrolment process to reduce data entry errors	125-126
Table 7.3	Factors in the enrolment process and their impact on various parameters	129-130
Table 7.4	Integrity of Aadhaar enrolments- respondents' views	130
Table 7.5	Responses on Inclusion of people with biometric exception	131
Table 7.6	Response in relation to updating of dynamic data	132
Table 7.7	Responses relating to standardisation of data collection processes	135
Table 7.8	Responses in relation to process of biometric collection	137
Table 7.9	Responses in relation to the issue of data leakage and data theft	139-140

Table 7.10	Responses in relation to the challenges of Aadhar enrollment	141
Table 7.11	Responses in relation to the issue of delay in receipt or non-receipt of Aadhar letter	143
Table 7.12	Normalised Scores of issues relating to enrolment eco-system.	144
Table 7.13	Responses to the issue of Safety and Security of Biometric Database	147
Table 7.14	Responses to the issue of Identity theft	148
Table 7.15	Responses to the issue of misuse of data by Enrolment Agencies	149
Table 7.16	Responses to the issue of security vulnerability due to PKI ownership status	150-151
Table 7.17	Summary of scores relating to issues of security and privacy	151
Table 7.18	Responses to the issue uniqueness of IDs	153
Table 7.19	Responses in relation to the issue of failure of authentication	155
Table 7.20	Responses the issue of robustness of biometric authentication	156
Table 7.21	Responses to the issue of safety and security of authentication on the Internet	157-158
Table 7.22	Normalised scores of issues relating to robustness of technology.	158
Table 7.23	Respondents' Comments around the issue of centralisation of Data by UIDAI	161-162
Table 7.24	Comments on the issue of Aadhaar and Privacy	164-165
Table 7.25	Respondents' Comments on the issue of de-enrolment facility	166-168
Table 7.26	Respondents' Comments on the issue of Private sector driven service delivery	169-170

Table 7.27	Respondents' Comments on the issue of Smart Cards by UIDAI	172-173
Table 7.28	Respondents' Comments on Aadhaar as a means of exclusion	174-175
Table 7.29	Respondents' Comments on Aadhaar and national security	177-178
Table 7.30	Respondents' Comments on lack of legal foundations of Aadhaar	179-180
Table 7.31	Respondents' Comments on duplication of work between UID and NPR	181-182
Table 7.32	Respondents' Comments on lack of legal foundations of Aadhaar	184-185
Table 8.1	Legal framework of Aadhaar contributing to second foundational principle of PbD Privacy as Default Setting	211
Table 8.2	Legal framework of Aadhaar contributing to second foundational principle of PbD Privacy Embedded in the Design	211
Table 8.3	Legal framework of Aadhaar contributing to sixth foundational principle of PbD Visibility and Transparency	213
Table 8.4	Legal framework of Aadhaar contributing to seventh foundational principle of PbD Keeping it user-centric	214
Table 10.1	Revisiting Research Objectives	254-255
Table A.4.1.1	Design Choices having an impact on Uniqueness	270-271
Table A.4.1.2	Execution Choices having an impact on Uniqueness	271
Table A.4.2.1	Design Choices having an impact on Inclusion	271
Table A.4.2.2	Execution Choices having an impact on Inclusion	271
Table A.4.3.1	Design Choices having an impact on Resident Convenience	272

Table A.4.3.2	Execution Choices having an impact on Resident Convenience	272
Table A.4.4.1	Design Choices having an impact on Cost optimization	272
Table A.4.4.2	Execution Choices having an impact on Cost optimization	272-273
Table A.4.5.1	Design Choices having an impact on Speed	273
Table A.4.5.2	Execution Choices having an impact on Speed	273
Table A.4.6.1	Design Choices having an impact on Scalability	273
Table A.4.6.2	Execution Choices having an impact on Scalability	274
Table A.4.7.1	Design Choices having an impact on Platform Approach	274
Table A.4.8.1	Design Choices having an impact on Future-proofing	274
Table A.4.8.2	Execution Choices having an impact on Future-proofing	274
Table A.4.9.1	Design Choices having an impact on Safety, Security and privacy	274-275
Table A.4.9.2	Execution Choices having an impact on Safety, Security and privacy	275

# List of Appendices

<b>Appendix Number</b>	<b>Description</b>	<b>Page Number</b>
Appendix A	Design and Execution Choices and their impact on over-arching goals of UID project	270-275
Appendix B	Pseudo Code of H-Soundex	276-277
Appendix C	A sample list of criticisms of Aadhaar and steps taken by UIDAI	278-281

## **List of Abbreviations Used**

AEPS	Aadhaar Enabled Payment Systems
ASA	Authentication Service Agencies
AUA	Authentication User Agencies
BPL	Below Poverty Line
BSP	Biometric Service Provider
CIDR	Central Identities Data Repository
CPU	Central Processing Unit
DBT	Direct Benefit Transfer
DDG	Deputy Director General
DDSV	Demographic Data Standards and Verification procedure
DL	Driving License
DMS	Document Management System
EGoM	Empowered Group of Ministers
EPIC	The Electronic Privacy Information Center
FI	Financial Inclusion
GDP	Gross Domestic Product
GoI	Government of India
ID	Identity document
IAS	Indian Administrative Services
ICO	Information Commissioner's Office (UK)

ICT	Information and Communication Technologies
KYC	Know Your Customer
KYR	Know Your Resident
LPG	Liquified Petroleum Gas
LRUR	Local Register of Usual Residents
MG-NREGA	Mahatma Gandhi National Rural Employment Guarantee Act
MHA	Ministry of Home Affairs
NADRA	National Database and Registration Authority (Pakistan)
NADRA	National Database and Registration Authority
NGO	Non-Governmental organization
NISG	National Institute of Smart Governance
NPR	National Population Register
NPR	National Population Register
NSAP	National Scholarships and Social Security Pensions
OTP	One Time Password
PAN	Permanent Account Number
PDS	public distribution system
PEC	Permanent Enrolment Centres
PET	Privacy Enhancing Technologies
PKI	Public Key Infrastructure
PMU	Project Management Unit
POA	Proof of Address

PoC	Proof of Concept
POI	Proof of Identity
QC	Quality Control
RASF	Remote Aadhaar Seeding Framework
RGI	Registrar General of India
RTI	Right to information
SFTP	Secure File Transfer Protocol
SRDH	State Resident Data Hub
SSN	Social Security Number
UIDAI	Unique Identification Authority of India
UNICEF	United Nations Children's Fund (previously United Nations International Children 's Emergency Fund)
VDM	Vendor Device Manager
WEF	World Economic Forum