

E-GOVERNANCE

**MISSION MODE PROJECT
(MMP)**

**CRIME AND CRIMINAL TRACKING NETWORK &
SYSTEMS**

RFP FOR SELECTION OF SYSTEM INTEGRATOR

ANNEXURE IV

DATA CAPTURE AND INTEGRATION

RELEASED BY:



Delhi Police

Government of Delhi

Table of Contents

1	INTRODUCTION	5
1.1	TYPES OF DATA	5
2	MASTER DATA	6
2.1	SOURCE OF MASTER DATA	7
2.1.1	<i>Organization Structure and Unit Hierarchy</i>	8
2.1.2	<i>Users at all locations</i>	9
2.1.3	<i>Acts and Sections of IPC/CrPC</i>	10
2.1.4	<i>Finger print details</i>	10
2.1.5	<i>Integration and Updation of Master Data</i>	10
2.1.6	<i>Integration Model</i>	11
3	LEGACY TRANSACTION DATA CAPTURE	11
3.1	METHODOLOGY FOR LEGACY TRANSACTION DATA CAPTURE:	12
3.1.1	<i>Recommendation: In view of the above the following is recommended</i>	16
3.1.2	<i>Suggested Digitization</i>	18
4	CONTENT/ DOCUMENT MANAGEMENT	18
4.1	DOCUMENT MANAGEMENT.....	18
4.1.1	<i>Methodology for Document Capture:</i>	19
4.2	INTEGRATION AND UPDATION OF CONTENT/DOCUMENT MANAGEMENT SYSTEM	22
5	QUALITY CONTROL OF DATA CAPTURE	23
5.1	MIGRATION STRATEGY.....	24
6	APPLICATION INTEGRATION	25
6.1	LIST OF APPLICATIONS AVAILABLE WITH DELHI POLICE DEPARTMENT.....	26
6.2	DETAILS OF APPLICATIONS RECOMMENDED TO BE INTEGRATED WITH CAS STATE.	28

List of Figures and Tables

Figure 1: Source of Master Data	8
Figure 2: Organization Structure and Unit Hierarchy	9
Figure 3: Users at all locations.....	9
Figure 4: Finger Print Details	10
Figure 5: Digitization of FIRs.....	13
Figure 6: Digitization of Data.....	16
Figure 7: Digitization Recommendation	17
Figure 8: Digitized Documents	19
Figure 9: Paper Based Documents.....	22

Annexure 4: Data Capture and Application Integration

1

1 Introduction

The purpose of this document is to develop an approach and formulate a strategy for legacy data capture for implementation of Crime and Criminal Tracking & Network System (CCTNS) for the State of Delhi. This project essentially involves implementation of an Enterprise Wide application to automate all the processes of Police. An implementation of this nature essentially requires certain data to be available without which it would not kick start. The aim of this document is to enable the availability of such data by virtue of the following

- (a) Identification of Data
- (b) Isolation of Data
- (c) Quantification of Data
- (d) Development of strategy for integration/capture of Data.

1.1 Types of Data

For any enterprise solution to be implemented, the following would be the categories of data that need to be captured and ported into the system.

- A. Master Data.
- B. Legacy transaction data to enable the system start up
- C. Data for the Content/Document Management System.

The above types of data would need a comprehensive deployment strategy to ensure smooth and integrated transition. The scope of this deployment strategy would include laying down the broad guidelines for the following:

- a) Preparation of a comprehensive data digitization plan.
- b) Data creation – Entries/Porting of FIRs, Investigation Details, Case Diaries, Prosecution Details, Verification Details Etc.
- c) Data migration & conversion – Including the conversion of existing computerized data into an appropriate format for the proposed new system.
- d) Testing of the conversion programs.
- e) Testing of the migrated data validating a particular reporting, auditing, or processing objectives as per the business rules.

2 Master Data

Master data comprises a set of the non-transactional data entities of an organization (also called reference data). This data forms the basis for designing and further carrying out transaction on the system. Management of such data is called Master Data Management (MDM). The main objective of MDM is providing processes for collecting, aggregating, matching, consolidating and quality-assuring, persisting and distributing such data throughout an organization to ensure consistency and control.

Issues

MDM seeks to ensure that an organization does not use multiple (potentially inconsistent) versions of the same master data in different parts of its operations, which can occur in large organizations. Other problems include issues with the quality of data, consistent classification and identification of data, and data-reconciliation issues.

One of the most common reasons some large organizations experience massive issues with MDM is integration. Two organizations which wish to integrate will typically create an entity with duplicate master data (since each likely had at least one master database of its own prior to the integration). Ideally, database administrators resolve such duplication in master data as part of the integration. In practice, however, reconciling several master data systems can present difficulties because of the dependencies that existing applications have on the master databases. As a result, more often than not the two systems do not fully merge, but remain separate, with a special reconciliation process defined that ensures consistency between the data stored in the two systems. Over time, however, as further integration occur, the problem multiplies, more and more master databases appear, and data-reconciliation processes become extremely complex, and consequently unmanageable and unreliable. Because of this trend, one can find organizations with 10, 15, or even as many as 100 separate, poorly-integrated master databases, which can cause serious operational problems in the areas of operational efficiency, decision-support, and regulatory compliance.

Solutions

Processes commonly seen in MDM solutions include source identification, data collection, data transformation, normalization, rule administration, error detection and correction, data consolidation, data storage, data distribution, and data governance.

The selection of entities considered for MDM depends somewhat on the nature of an organization. In this case, MDM may apply to such entities as Police Units, Unit Entitlement, Personnel Data and list of offences/Acts/IPC. MDM processes identify the sources from which to collect descriptions of these entities. In the course of transformation and normalization, administrators adapt descriptions to conform to standard formats and data domains, making it possible to remove duplicate instances of any entity. Such processes generally result in an organizational MDM repository, from which all requests for a certain entity instance produce the same description, irrespective of the originating sources and the requesting destinations.

2.1 Source of Master Data

Police Organizations have several Police Stations and other units spread all across the state. Delhi has been trifurcated in three ranges and further into 11 Districts for the purpose of maintaining law & order. Each range is headed by an officer of the rank of Joint Commissioner of Police and the District is headed by a Deputy Commissioner of Police.

There are 11 districts in Delhi Police. The overall supervision of the Police stations/Police posts within the Distt. rests with respective Deputy Commissioners of Police. Prevention, investigation and detection of crime and maintenance of law & order are the important functions of police. The duty of the Officer in charge of Police Station is to prevent & detect crime and maintain law & order within his/her jurisdiction. The Deputy Commissioner of Police of District supervises the sub-divisions, Police Stations/Police Posts under his/her control

Currently the level of automation in the state police includes the implementation of CIPA (Common Integrated Police Application) which has been developed by NIC on linux (Knoppix) platform using Postgre as the standard database, along with various other applications like zipnet, dossier etc. The aim of CIPA was to automate the police functioning as far as registration, investigation and prosecution processes are concerned.

Following are the Master Data that need to be captured/ported into the CCTNS system for it to kick start :

- (a) Organization Structure of the complete organization
- (b) Unit wise hierarchy
- (c) Details of users at all locations
- (d) Acts and Sections of IPC/CRPC
- (e) Finger Print Details

In addition to this master data, several other master tables will have to be created and corresponding data identified and entered for each module of the application while designing the system in whole. In addition to the above mentioned master data, masters relating to the following data need to be created:

- (a) Delhi Police Act 1978
- (b) Punjab Police Rules
- (c) Juvenile Justice Act
- (d) Indian Railway Act
- (e) DMRC Act

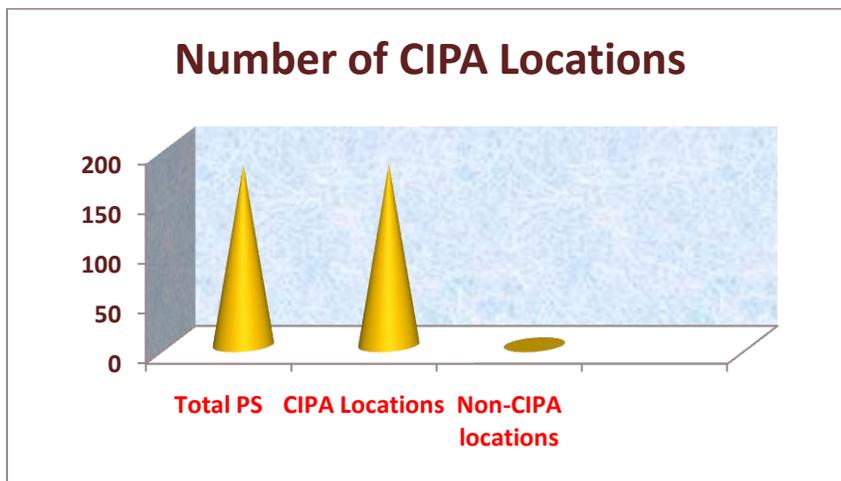


Figure 1: Source of Master Data

2.1.1 Organization Structure and Unit Hierarchy

These masters would map the complete organization structure of Delhi police. This data would primarily be needed to denote the dependencies/reporting of various police stations on their respective HQ. **This data is to be captured afresh since no such database exists in a structured format.**

Sr No.	DATA - Delhi Police	
1)	Police Stations-	
	Total number of police stations	181
2)	Higher Offices	70
3)	SDPO	60
4)	Equivalent to SDPO (within District- ACP office)	55
5)	Equivalent to SDPO (treated like Districts - ACP office)	29
6)	Districts	11
7)	Equivalent to Districts (Crime & Railways, IGI Airport, EoW, Special Cell, Special Police Station for Women & Child)	5
8)	Ranges	3
9)	Equivalent to Ranges (Crime & Railways, IGI Airport, EoW, Special Cell, Special Police Station for Women & Child)	5
10)	PHQ- Large	1
11)	Police HQ Medium (Traffic & PCR)	2
12)	SCRB	1
13)	District control rooms	16
14)	FSL	1
15)	FPB	1
16)	Other Details -	
17)	Number of RTs/PTCs/DTCs (Existing)	4

18)	Number of RTSS/PTCs/DTCs (Proposed)	12
Total Number of Units		457

Figure 2: Organization Structure and Unit Hierarchy**2.1.2 Users at all locations**

These masters would map the user set of the system. User set refers to the complete set of the users who would log into the system to carry our various functions. The no. of users would differ at each level and location. However, to arrive at fair estimates, the following assumptions have been made:

S/No	Entity	Estimated no of Users Per Location	Total locations	Total Users	Estimated no of fields	Estimated no of records
1	Police Stations	15(1* SHO, 7* Investigation Officers, 2* Computer Operators, 3* HM, 2* records Mgt)	181	2760	20	2760
2	Higher Offices	10(5 Officer, 2* Operator, 3* Records)	74	740	20	740
3	SDPO	4 (1 * ACP , 1 * Computer Operator, 2 * Record Management)	144	576	20	576
4	Range Jt. CP	12 (2 * Officers , 4 * Crime Sec , 2 * DSB , 2 * Staff , 2 * PA)	8	96	20	96
5	PHQ (medium)	14* 6 (6 users per department)	2	168	20	168
6	SCRB	10 * 6 (6 users per dept)	1	60	20	60
7	Commissionrates (PHQ- Large)	10*6(6 Users per dept)	10(Separate records for each Dept, 10 Depts)	60	20	60
8	RTCs /PTC s	5 users	16	80	20	80
9	District offices	12 (2 * Officers , 4 * Crime Sec , 2 * DSB , 2 * Staff , 2 * PA)	16	192	20	192
10	District control rooms	5	16	80	20	80
11	FSL & FPB	5	2	10	20	10
Total						4822

Figure 3: Users at all locations

Apart from the above mentioned data, No, Rank, Name, Seniority, Date of Enrolment etc can be ported from the PIS system and the rest needs to be captured afresh. The data that needs to be captured afresh would essentially be related to the posting details, appointment details etc.

2.1.3 Acts and Sections of IPC/CrPC

These masters would map the complete legal frame work under which the police work. This would include the following

- (a) Indian Penal Code
- (b) Code of Criminal Procedures
- (c) Delhi Police Act
- (d) Punjab Police Rules
- (e) Motor Vehicle Act
- (f) Local and special Acts of centre and States

Most of the master data related to all the above mentioned has been captured as a part of CIPA implementation, hence it is identified as the authentic source and hence can be ported from the current implementation of CIPA.

2.1.4 Finger print details

These masters would be required for the Fingerprint module of CCTNS application. At the state Finger Print Bureau, AFIS (Automated finger print identification system) has been implemented which has a database of around three lakh fingerprint sets. This data is primarily related to all people arrested and all repeat offenders. AFIS has been identified as the authentic source of fingerprint data and hence will be integrated with the CAS (State) Application.

S/No	Table	Fields	No of Records
1	Finger Information(AFIS)	Print 40	Approx 3 Lac

Figure 4: Finger Print Details

2.1.5 Integration and Updation of Master Data

In the above section, the sources of master data have been identified and the parts that need to be ported into the CAS (State) have been specified. Once the initial porting is over, the master data needs to get updated regularly in order to keep it current. CAS (State) system would be hosted centrally through a Central Data Centre (Co located with the State Data Centre). This essentially means that the master data would reside centrally at the central database and would service all the users from this central location. The issues that need to be addressed while designing the system would be

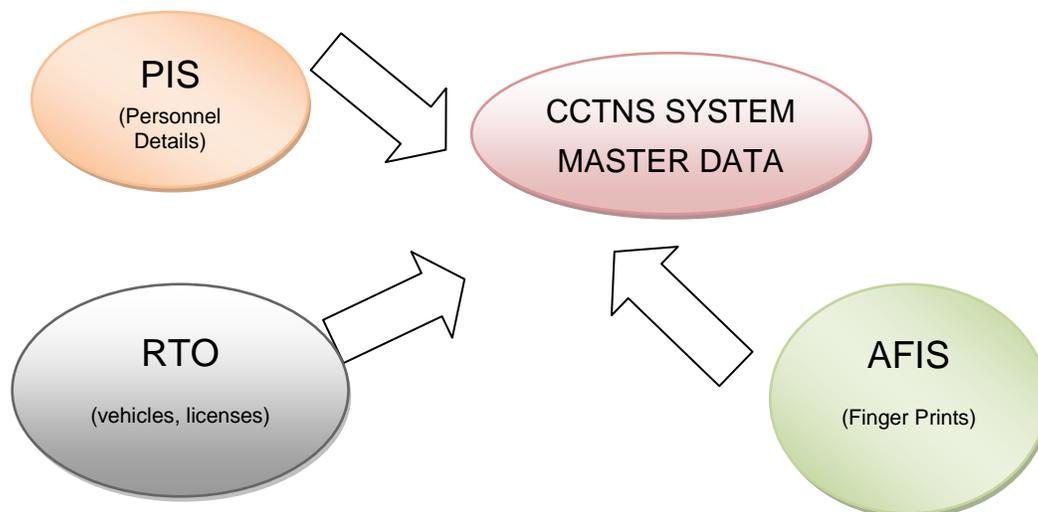
- (a) **Integration:** Since integration with the external systems (PIS, RTO, AFIS) need to be dynamic, it is important that the integration happens at the database level. In essence the CAS (State) database needs to be designed based on the data structures of the external systems. This would enable ease

of porting the data into the CAS (State) systems, where required. In addition, it is pertinent to note that the integration would be substantiated by a live link between CAS (State) system and the external systems through various security levels and access rights.

- (b) **Updation:** Any change/update in the master data in the external systems has to be reflected in the CAS (State) system too. This process of updation can be scheduled on occurrence or at a given time interval. This would ensure that the CAS (State) system is current and updated.

2.1.6 Integration Model

CCTNS Master Data will include organization structure, unit hierarchy, appointment details, posting details, etc of all the police personnel. It will collate all relevant data under one umbrella to ease police functioning.



3 Legacy Transaction Data Capture

In addition to all the types of masters discussed above, there is another important type of data that needs to be captured in order to make the implementation of CCTNS system effective. This is the legacy transaction data that gives the history of transactions especially in the terms of FIR, Investigation Reports etc. Every activity in the envisaged CAS application involves a transaction in complaint registration, Investigation, Prosecution etc. As FAR as CIPA implementation is concerned FIR registration, Investigation and Prosecution functions have been automated. However all the other functions are manual. These manual transactions in the current system are recorded in ledgers, registers and files. Certain activities which create transaction data are listed below

- (a) Registration of FIR
- (b) Registration NCR
- (c) Registration of People(Foreigners, Proclaimed Offenders, Prisoners, Arms License)
- (d) Verification(Passport, Service, Character)
- (e) Investigation(Crime Details, Arrest/ surrender, seizure, chargesheet/FR)

- (f) Prosecution(Appeal, Bail, Movements, Reopening and Reassigning, Summons/ Warrants, result of the case)
- (g) FSL(Inward receipt and Outward receipt)
- (h) Missing person

The capture of this legacy transaction data is the most difficult part of data capture effort in case of any implementation, because this data needs to be identified, cleansed and ported. The activities involved in the capture of this kind of data are as follows

- (a) Identification and Isolation of Digitized Data
- (b) Identification and isolation from paper based forms/registers/ledgers
- (c) Formatting based on the database structure
- (d) Cleansing and validation
- (e) Entry into the database through forms
- (f) Testing

3.1 Methodology for legacy Transaction Data Capture:

Digitization of FIRs: All police stations have implemented CIPA in Delhi, i.e. a total of 181

have implemented CIPA. As part of CIPA implementation FIR registration, and some parts of Investigation and Prosecution processes has been automated, i.e. all the relevant data regarding FIR registration, and parts of Investigation and Prosecution is available in the digitized form. However, the quality of the data in CIPA is not yet tested and the reliability of this data needs to be checked.

S.N.	Offices	Number of FIR Records (All types) (cognizable offences only)		Remarks
		1 Jan 1999 to 31 Dec 2007	1 Jan 2008 to 31 Dec 2008	
1	South East	79202	6242	CIPA
2	South	67966	4812	CIPA
3	West	94940	5317	CIPA
4	South West	63667	4072	CIPA
5	New Delhi	25274	1857	CIPA
6	North East	57860	4680	CIPA
7	East	59067	5933	CIPA
8	Outer	65452	5227	CIPA
9	North West	57691	3959	CIPA
10	North	50639	3518	CIPA
11	Central	61684	3527	CIPA

12	IGI Airport	6078	504	CIPA
13	Crime & Railways	14315	908	CIPA
14	Narcotics	785	112	CIPA
15	E.O.W.	0	216	CIPA
16	Special Cell	698	65	CIPA
17	C.A.W.	0	91	CIPA
Total		121358	24983	146341

Figure 5: Digitization of FIRs

From the above it is evident that the no of FIR digitized as part of CIPA Implementation for the last 10 years is approximately 1.46 lakh. However the records in CIPA prior to the year 2006 are not complete and hence will have to be digitized for CAS. The total numbers of records from Jan 1999 to Dec 2005 are 94389 (calculated on a yearly average basis). In order to arrive at a realistic figure it is important to take a standard sample which is most common, standard and most used. In this case we have considered a standard FIR as a template for assessing this data.

- (a) No of fields in a standard FIR form IIF-1 : 69 (Form 1 + Attachment)
- (b) Average no. of characters in each field: 20(Some of the entries like address of complainant, details of accused, reasons for delay etc are large but a majority of fields need small entries like date, Name etc.
- (c) Each record size : $69 * 20 = 1380$ Characters
- (d) After appending an additional 3000 Characters for the FIR detail section, total number of characters per FIR record will be 4380 Characters.
- (e) Total no of characters that need to be digitized now for NON CIPA is $4380 * 94389 : 413$ MB

Digitization of NCR:

After detailed analysis it is assumed that each police station receives about 2000 NCR per year. I.e. for a total of 184 police stations (the total no of NCR= $8000(\text{average}) * 184 = 1472,000$

I.e. for the last 10 years it can be assumed that about 14720000 NCR registrations. Hence taking the above estimates of 50 fields per form and 20 characters per form, the same would get translated into

approximately 14.7 GB of Data that needs to be digitized

Digitization of Complaints:

After detailed analysis it is assumed that each police station receives about 6000 complaints per year. I.e. for a total of 184 police stations (the total no of NCR= $6000(\text{average}) * 184 = 1104,000$

i.e. for the last 10 years, it can be assumed that about 11040000 complaint registrations. Hence taking the above estimates of 50 fields per form and 20 characters per form, the same would get translated into

Approximately 11 GB of Data that needs to be digitized.

Registration of People (Foreigners, Proclaimed Offenders, Prisoners, Arms License)

After detailed analysis it is assumed that each police station receives about 100 registration requests of various kinds that include foreigners, proclaimed offenders, prisoners, arms license etc per year. I.e. for a total of 184 police stations the total no of people registration= $100(\text{average}) * 184 = 18400$

i.e., for the last 10 years it can be assumed that about 184,000 people registrations. Hence taking the above estimates of 50 fields per form and 20 characters per form, the same would get translated into

Approximately 184 MB of Data that needs to be digitized.

Verification (Passport, Service, Character)

After detailed analysis it is assumed that each police station receives about 3840 verification requests of various kinds that include passport, character etc per year, i.e., for a total of 184 police stations, the total no of verifications= $3840(\text{average}) * 184 = 7,06,560$

i.e., for the last 10 years it can be assumed that about 70,65,600 verifications. Hence taking the above estimates of 50 fields per form and 20 characters per field, the same would get translated into

Approximately 7 GB of Data that needs to be digitized.

Investigations (Arrest, seizure, FSL, Surrender)

After detailed analysis it is assumed that each police station receives about 500 investigations of various kinds that include arrest, seizure, surrender etc per year. I.e. for a total of 184 police stations the total no of investigations= $500(\text{average}) * 184 = 92000$ i.e. for the last 10 years it can be assumed that about 320000 investigations. Hence, taking the above estimates of 50 fields per form and 20 characters per field, the same would get translated into

Approximately 920 MB of Data that needs to be digitized.

Prosecution (Appeal, Bail, Movements, Reopening and Reassigning, Summons/ Warrants)

After detailed analysis it is assumed that each police station receives about 50 prosecutions of various kinds that include appeal, bail etc per year, i.e. for a total of 184 police stations the total no of Prosecutions= $50(\text{average}) * 184 = 9200$

i.e., for the last 10 years, it can be assumed that about 92000 prosecutions. Hence, taking the above estimates of 50 fields per form and 20 characters per field, the same would get translated into

Approximately 920 MB of Data that needs to be digitized.

Traffic (Challaning)

After detailed analysis it is assumed that traffic police receives about 5, 40,000 cases per year.

i.e., for the last 10 years, it can be assumed that about 54, 00,000 prosecutions. Hence, taking the estimates of 10 fields per form, and 20 characters per form, the same would get translated into

Approximately 5.4 GB of Data that needs to be digitized.

FSL (Inward receipt and Outward receipt)

After detailed analysis it is assumed that FSL receives about 5740 requests per year, i.e., for the last 10 years, it can be assumed that about 57,400 cases. Hence, taking the above estimates of 50 fields per form and 20 characters per form, the same would get translated into

Approximately 57.4 MB of Data that needs to be digitized.

Fingerprint Bureau

After detailed analysis it is assumed that Fingerprint Bureau receives about 2180 (500 photographs and 1680 fingerprint) requests per year.

i.e. for the last 10 years, it can be assumed that about 21,800 cases. Hence, taking the above estimates of 50 fields per form and 20 characters per form, the same would get translated into

Approximately 21 MB of Data that needs to be digitized.

EOW

After detailed analysis, it is assumed that EOW receives about 1200 requests per year. i.e., for the last 10 years, it can be assumed that about 12,000 cases. Hence, taking the above estimates of 50 fields per form and 20 characters per form, the same would get translated into

Approximately 12 MB of Data that needs to be digitized.

RTI

After detailed analysis, it is assumed that each Police station receives about 730 RTI requests per year. i.e., for the last 3 years, it can be assumed that about 4, 02,960 cases. Hence, taking the above estimates of 50 fields per form and 20 characters per form, the same would get translated into

Approximately 4 MB of Data that needs to be digitized.

S/No	Activity	Assumption	Data Requirement	Digitization	Remark
1.	FIR Registration (CIPA Locations)	Actuals	413 MB		To be Digitized as part of CIPA

2.	NCR	8000 per PS per year	14.7 GB	To be digitized
3.	Verification	3840 per police station	7 GB	To be digitized
4.	Investigation	200 Per PS per year	920 MB	To be digitized
5.	Prosecution			To be digitized
6.	FSL	5740 requests per year	57.4 MB	To be Digitized
7.	EOW	12000 requests per year	12 MB	To be Digitized
8.	Traffic	5,40,000 per year (Entire Delhi)	1.08 GB	Not to be digitized
9.	Complaint	6000 per PS per year	11 GB	Not to be digitized
10.	People Registration	100 Per PS per year	184 MB	Not To be digitized
11.	Fingerprint (Photo section)	500 requests per year	21 MB	Not To be Digitized
12.	Fingerprint (Fingerprint section)	1680 requests per year		Not To be Digitized
13.	RTI	730 per PS per year	4 MB	Not To be Digitized
Grand Total			35.38 GB	Total Data
Total (to be digitized)			22.5 GB	To be digitized

Figure 6: Digitization of Data

The SI will be required to calculate the total effort required for the digitization task

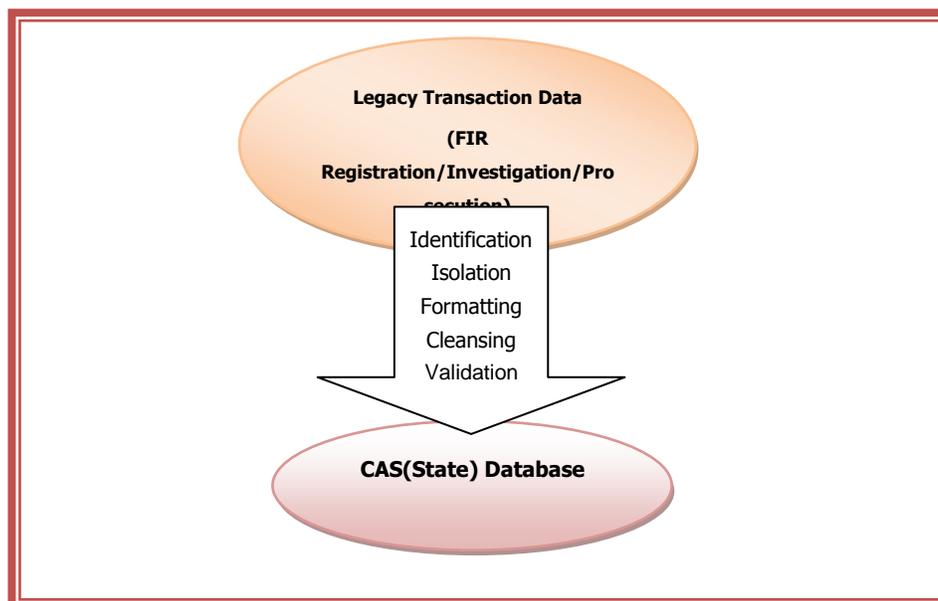
3.1.1 Recommendation: In view of the above the following is recommended

- (a) **FIR Registration:** This Data is critical and has legal implications hence it is recommended that all FIRs not digitized as part of CIPA be ported into the new system along with the metadata.
- (b) **Investigation/Prosecution/FSL/EOW:** This Data has limited legal implications. Hence it is recommended that all **NCR/Complaints/Investigation/Prosecution** for the last 1 year be digitized to update the system. The rest can be maintained in manual documents till their legal validity is over.
- (c) **Registration of People/ Verification/ Traffic/ Fingerprint bureau//RTI NCR Registration/Complaint:** This Data has lesser implications. Hence it is recommended that no legacy data be digitized. The legacy data can be maintained in manual documents till their legal validity is over.

Based on the above recommendations the amended work sheet is as follows

S/No	Activity	Assumption	Data Requirement	Digitization	Remark
1.	FIR Registration (CIPA Locations)	Actuals	413 MB		To be digitized as part of CIPA
2.	Investigation		92 MB		To be digitized
3.	Prosecution				To be digitized
4.	FSL	5740 requests per year	5.74 MB		To be Digitized
5.	EOW	12000 requests per year	1.2 MB		To be Digitized
6.	Traffic	5,40,000 per year (Entire Delhi)	0		Not to be digitized
7.	Complaint	6000	0 GB		
8.	NCR	8000 per PS per year	0GB		Not to be digitized
9.	Verification	3840 per police station	0		Not To be digitized
10.	People Registration		0		Not To be digitized
11.	Fingerprint (Photo section)	500 requests per year	0		Not to be Digitized
12.	Fingerprint (Fingerprint section)	1680 requests per year			Not to be Digitized
13.	RTI	730 per PS per year	0		Not to be Digitized
Total (to be digitized)			512.94 MB		To be digitized

Figure 7: Digitization Recommendation



3.1.2 Suggested Digitization

Following is a list of registers which are permanent registers. Hence, it is recommended to digitize the documents since inception

S/No	Register	No of Columns	No of Pages covering last one year transactions
1	VCNB	20	200
2	Deserter register	22	10
3	Absconder / Proclaimed Offender register	22	10
4	Arms license Register	26	10
5	Habitual offender register	22	10
6	Index register	10	120(10 pages per month)
7	Unnatural death register	28	10
8	Crime Index Register	22	120(10 pages per month)
		Total	490

4 Content/ Document Management

4.1 Document Management

As part of the day to day functioning of each police station, a large no of documents are generated by each and every functionary. Currently several of these documents are paper based. It would be pertinent to mention here that all police stations are computerized to an extent where in all the documents are generated using computer applications such as Word star, MS Word and Adobe Pdf. In addition to these digitized documents, there still are some paper based documents whose digitized versions may not be available with the police stations.

The documents which are available in the digitized form would include

- (a) Everyday correspondence
- (b) Reports and returns

The documents which may not be available in digitized form at Police units would be:

- (a) In coming correspondence
- (b) Policy Letters
- (c) Police Manuals

4.1.1 Methodology for Document Capture:

The proposed application would have a robust document management system which would primarily be indexed repository of all the documents to be generated and hosted by the application.

4.1.1.1 Digitized Documents: These documents would be directly ported into the document management system after indexing. These documents would be segregated and classified into folders based on their role and index. The total approximate size of such documents is as given below

Assumption: Size of a word document (40 lines per page) = 25 Kb per page

Size of a pdf document (40 lines per page) = 50 Kb per page

S/No	Entity
1.	Commissionerate (Incl'd PHQ)
2.	District Office
3.	SDPO
4.	Police Stations (including IGI, Metro, Railways, etc.)
5.	SCRB
6.	Women And Child Cell
7.	Special Cell

Figure 8: Digitized Documents

4.1.1.2 Paper Based Documents:

The functioning of most police stations happens through these registers. As part of the CIPA implementation, only FIR register has been automated, i.e., FIRs are registered on the system and the register is maintained by the system. The additional 22 registers are manually maintained. Any automation project requires these registers to be electronically maintained. Since the above mentioned data digitization activity will ensure that most of the above mentioned registers are automated, hence it is recommended that the following registers be scanned and captured as part of the content/document management system.

S.No.	Nature of Record	No of Columns	No of Pages covering last one year transactions per Police Station	Approx. total no of Pages covering last one year transactions (entire Delhi)*	Remarks
1.	First Information Report (FIR) maintained as per PPR-22.47 (Register No.1)	15	NA	74,949	Digitized in CIPA

S.No.	Nature of Record	No of Columns	No of Pages covering last one year transactions per Police Station	Approx. total no of Pages covering last one year transactions (entire Delhi)*	Remarks
2.	Daily Diary maintained as per PPR 22.48 (Register No.II)	4	10,000	18,40,000	To be Digitized
3.	All Standing Orders as per PPR 22.53 (Register No.III)	NA	NA	4,500	Available in digitized form on IntraDP
4.	Correspondence register maintained as per PPR 22.55 (Register No.V)	6	1320	2,42,880	To be Digitized
5.	Inquest register maintained as per PPR 22.56 (Register No. VI)	22	10	1,840	To be Digitized
6.	Crime Record Register maintained as per PPR-22.59 and PPR-22.60 (Register No.IX)	8	250	46,000	To be Digitized
7.	Registers of Information Sheets maintained as per PPR-22.63 (Registers No. XII and XII-A)	9	300	55,200	To be Digitized
8.	Register for Arms Act licenses, licenses under Excise laws, licenses under Explosives act, licenses under petroleum act, licenses under poisons act and sarais under Sarais Act maintained as per PPR-22.68 (Register No. XVII)	9	30	5,520	To be Digitized
9.	Arms & Ammunition Receipt book maintained as per PPR-22.69 (Register No. XVIII)	6	5	920	To be Digitized
10.	Case Property Register No.XIX maintained as per PPR-22.70.	8	250	46,000	To be Digitized
11.	Road Certificate maintained as per PPR-22.72 (Register No. XXI)	7	1000	1,84,000	To be Digitized
12.	Copies of all Police gazette, Criminal intelligence gazette and all orders maintained as per PPR-22.74 (Register No. XXIII)	NA	NA	NA	Already digitized with CRO branch

S.No.	Nature of Record	No of Columns	No of Pages covering last one year transactions per Police Station	Approx. total no of Pages covering last one year transactions (entire Delhi)*	Remarks
13.	Copies of all Police rules maintained as per PPR-22.76 (Register No. XXIV)	NA	NA	NA	Already digitized with CRO branch
14.	Missing Person Register	14	10	1,840	Digitized in CIPA
15.	MLC Register	8	100	18,400	Digitized in CIPA
16.	PCR Call / DD entry register	9	400	73,600	To be Digitized
17.	Arrested persons register	14	100	18,400	To be Digitized
18.	Bail Register	...	30	5520	To be Digitized
19.	Surety register	8	30	5520	To be Digitized
20.	Complaint register		375 (1500 entries/4 entry per page)	69000	To be Digitized
21.	Pairvi of heinous cases				To be Digitized
22.	Ruffian register part-1		5 (25/5)	920	To be Digitized
23.	Ruffian register part-2		30 (150/5)	5520	To be Digitized
24.	Ruffian register part-3		1	184	To be Digitized
25.	Ruffian register part-4		1	184	To be Digitized
26.	Servant verification register	8	100	18400	To be Digitized
27.	Tenant verification register	8	100	18400	To be Digitized
28.	Kalandra register	11	200	12,8,800	To be Digitized
29.	Case file / challan movement register	9	24,983	45,96,872	To be Digitized
Grand Total				5286649	

S.No.	Nature of Record	No of Columns	No of Pages covering last one year transactions per Police Station	Approx. total no of Pages covering last one year transactions (entire Delhi)*	Remarks
Total				5188800	

Figure 9: Paper Based Documents

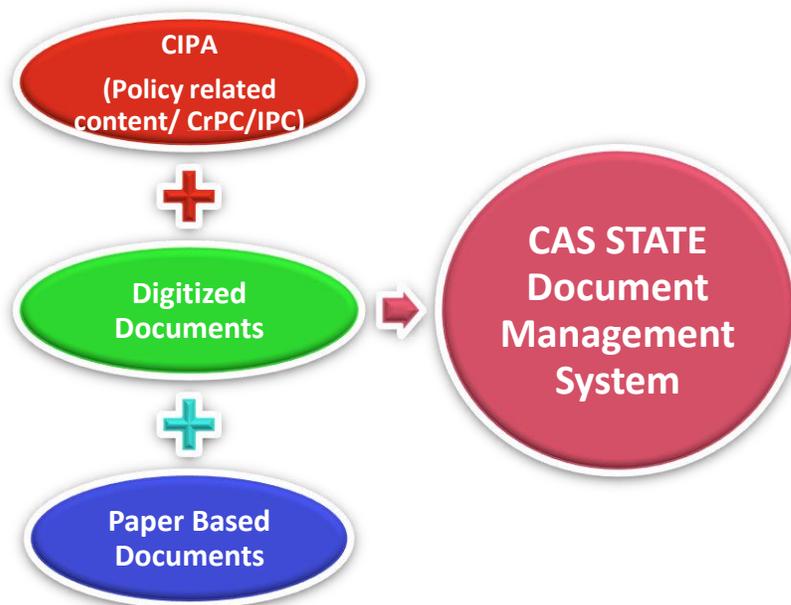
An approximate **51, 88,800** number of pages need to be scanned and uploaded to the content /document management system for all Police Stations of Delhi Police. Assuming each scanned page to be 40Kb, to **total content to be scanned will be 207 GB**

4.1.1.3 Meta Data: Meta data relates to all the attribute data for the documents which identify the documents with respect to the following

- (a) Date of creation/modification
- (b) Author
- (c) Date of routing/approval
- (d) Subject
- (e) Key content
- (f) References
- (g) Key words for search

This attribute data also needs to be captured as part of the data capture effort for the documents.

4.2 Integration and Updation of Content/Document Management System



5 Quality Control of Data Capture

Challenge of data conversion / migration is the coordination between three normally disparate groups

- ▶ Functional (From Business)
- ▶ Legacy functional
- ▶ Legacy technical

NISG & PwC strongly recommends that the project assigns a data conversion manager.

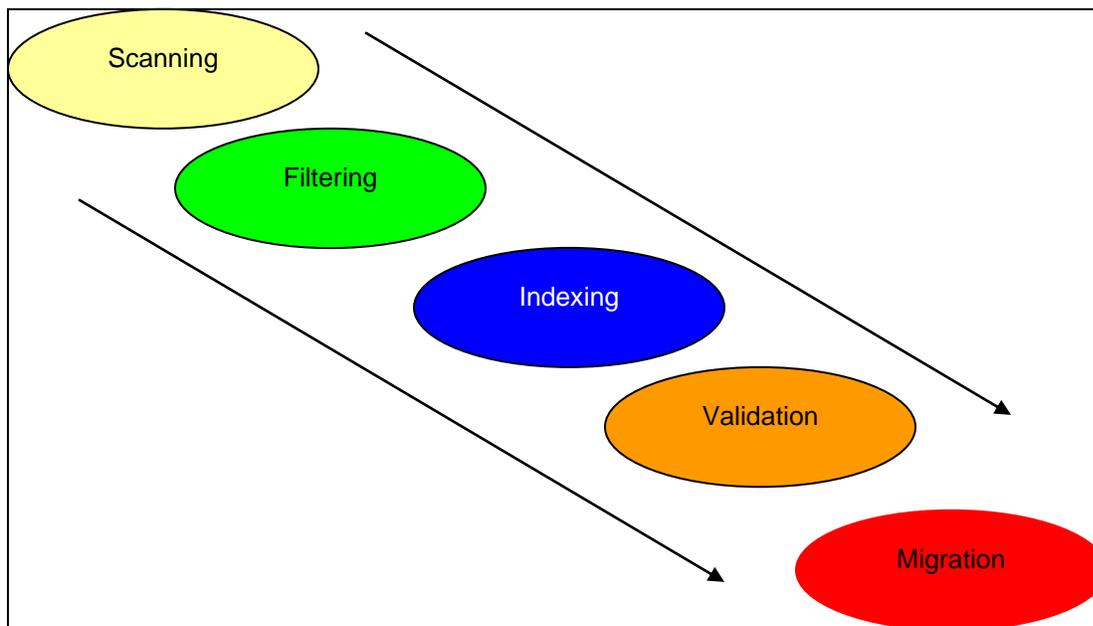
Data conversion process will ensure that there is a focus on conversion requirements from the start of the project. Overall conversion strategy for the project determines the approach for data conversion and the tools and methods to be used for conversion. Data requirements for the test database, conference pilot and the Production database will be met by this process.

- a) The conversion process will include designing and building conversion programs and the actual data conversion from the legacy systems. The conversion process will ensure that the team is prepared in advance with extraction scripts for the legacy systems and the upload scripts for the new applications.
- b) Data conversion is accomplished through manual and/or automated means and includes the creation of any data needed for new data records. It enables the new system to accurately reflect the current status of organizational data at the start of production. The term "data conversion" is used in the method in a wider sense than the dictionary definition. It is defined to include four separate elements:
 - Purging/cleansing of existing data.
 - Reconciliation/balancing of the old data and the old system/new system.
 - Creation of new/additional data.
 - Conversion of the old data to the new system to create master file data and opening balance data (where appropriate) for the new system.
- c) The conversion of data from an old system to a new system usually requires the creation of a conversion system. The conversion system can be manual, automated or both. Logically, conversion is taking data from one system and moving it to another system at a point in time.
- d) Conversion may translate the format of the data but not update it and Conversion programs may be run more than once or multiple times with each execution converting a subset of data, such as one calendar month.
- e) May be run multiple times for parallel running, phased or staged implementation where progressive data conversion is required, and may be run multiple times until reconciliation shows a valid conversion.
- f) System development support may be required to create a data conversion system. In addition to the normal problems of system development, data conversion has several unique requirements.
- g) Data may need to be created if it does not exist. Current data may need to be purged, corrected or verified. The existing system (manual or automated) may need to be corrected or balanced within itself before conversion can be commenced or completed;

- h) Timing of the data conversion process may vary, e.g., data cleansing or creation may precede other tasks by several months.
- i) The following is a description of the various tasks that make up the data conversion activity:
- a. Analyze Data Conversion - Define and analyze data conversion entities, both on the legacy systems and on the receiving ERP Application.
 - b. Design/Build Conversion Programs - Design, construct and unit test all conversion programs.
 - c. Implement Feedback from EP/Integration Testing - Update design and code for conversion entities, based on input from system prototyping and testing
 - d. Prepare for Cutover - Make final conversion modifications to prepare for system cutover.

5.1 Migration Strategy

The data capture effort of legacy data can be divided into the following phases:



- **Scanning:** The physical scanning of the files and documents would be the first activity in this process. Once all the relevant documents of a case file or a complaint file are scanned they will be stored in a standard format clearly indicating its content.
- **Filtering:** Each such file once scanned will be filtered for any irrelevant content. During this phase it will also be ensured that the quality and sequence of the document is in conformity to the original. The filtered document will be saved in the standard format clearly indicating its parameters like Case/ Complaint No, Date etc.
- **Indexing:** The case file once filtered will be indexed by entering its parameters to the database. These parameters will be decided during the design phase of the application. Once indexed and logged the completed case file is ready to be ported onto the system and made live.

- **Validation:** Before each case is ported onto the system, validation is an important activity that needs to be carried out for ensuring consistency and integrity. During this phase the activities in all the above phases will be validated and verified.
- **Migration:** The validated case files are finally ported into the system and then data is live for use.

6 Application Integration

In addition to the Configuration, Customization and Extension (New Modules) of CAS (State), it is imperative for the entire solution to get integrated with the CAS (Center) and External Agencies. System Integrator will be required to integrate the complete set of solution with CAS (Center) and create interfaces for further integration with External Agencies.

6.1 List of applications available with Delhi Police Department

S.No	Application Software Name	Functionalities	Area/ police stations covered	Platform /database	Current status	Other details
1	Zonal Integrated Police Network (ZIPNET)	Coordination of Lost & Found vehicles, Mobiles & Kidnapped , Missing, Dead Persons etc.	CRO , 175 police stations	Red Hat Linux My SQL Db Front end	Working	In Collaboration with 7 neighboring states
2	MACT Application	Database of all the vehicles and their owners	All police stations	.NET Microsoft SQL Db	Working	
3	Crime Criminal Information System (CCIS)	Records of criminal Case Entry	Internal application at CRO	Client -Server Oracle 10g Db	Working	LAN based application.
4	Online Dossier System	Digitization of FIRs, Criminal data (features, photographs etc)	Internal Application, CRO	Oracle Db	Working	LAN based application
5	GIS-GPS based Automatic Vehicle Tracking System (AVTS) for PCR vans	A GIS based application	Internal	CAD based system	Working	GIS system which will guide the Police vans.
6	Automated Fingerprint Identification System (AFIS)	Collection and Coordination of Criminal's Finger Prints	Internal	FACTS (Finger print Analysis and Criminal Tracking System	Working	
7	Traffic SMS alert	SMS alert Service for diversion, blockage etc of traffic	Internal Application (Traffic Control room)	Local service provider	Working	
8	Traffic Mapping system	A GIS based application for directing traffic police personnel	Internal	Arc info and Auto CAD	Working	
9	Automated Vehicle Information System(AVIS)for traffic vehicles	A GIS based application to track the location of control vehicles	Internal Application	DELL Servers Linux OS SQL Db	Working	
10	Traffic Monitoring System	For Smooth movement of vehicles	36 Cameras are placed at 9 locations	Supported by application for	Working	

		To track vehicles Moving against law Crossing the stopline Defective number plates has Videp Analytics Without helmet Without seatbelt (with some difficulty)	(26 Static and 10 PTZ Cameras)	taking snapshots and storing it in db (Oracle 10g)		
11	Teleform Verifier	For reading challans and digitizing it instantly	Combines ownership and Offense data	MS ACCESS / VB based application	Working	Software was supplied by SIEMENS
12	PC Crash	Provides a facility to simulate an accident by creating a virtual accident scenario	Assists in Accident research	Proprietary Software	Working	
13	Web Portal at HQ (Intra DP) www.delhipolice.nic.in	Details of Personnel Complaint Monitoring Court Cases Telephone Directory Summon Warrants Standing Orders / Circulars	Exchange Server SQL Database ASP / ASP.NET Cluster Services Domain Controllers	6 servers XEON based SAN-460 GB 80GB/320 GB HDDs		
14	delhipoliceqac.in	For allotment of quarters	.NET /SQL			
15	Transfer Process Application	For managing the transfer process. Routine processes are automated	Internally Developed VB / SQL			
16	PCR100 number facility	60 call receiving stations for taking calls from public and forwarding it to respective police stations	e-forms are generated and moved across			Others Child Helpline Women Helpline Senior Citizens helpline etc.
17	C4I	Working	Application hosted in the data center at PHQ		Internal	Not Part of CAS application. This application will be independent of CCTNS
18	Inventory Control System	Working	Application hosted in the data center at PHQ		Internal	Not Part of CAS application. This application will be independent of CCTNS

6.2 Details of Applications Recommended to be integrated with CAS STATE.

S.No	Application Software Name	Current status	Database/ Records	Integration with CCTNS	Internal / Public	Nature of integration with CCTNS
1	Zonal Integrated Police Network (ZIPNET)	Working	3.5 Lac records a total of 8GB data	Yes	Internal /Public	This functionality is a part of CAS Services. The existing database will be migrated
2	MACT Application	Working	7,500 records per year	Yes	Internal	This functionality is a part of CAS Services. The existing database will be migrated
3	Crime Criminal Information System (CCIS)	Working	8 Lakh records of 20 GB data. Located at CRO office	No	Internal	Will not be applicable post CAS implementation. Data will be migrated
4	Online Dossier System	Working	60000 records of about 5 GB data. Located at CRO Office	No	Internal	Will not be applicable post CAS implementation. Data will be migrated
5	GIS-GPS based Automatic Vehicle Tracking System (AVTS) for PCR vans	Working	Application hosted at Traffic HQ	No	Internal	This is an independent system. Cannot be integrated with CCTNS at this point
6	Automated Fingerprint Identification System (AFIS)	Working	The database contains 3 lakhs records of finger prints	Yes	Internal	CAS has a similar functionality and the data will be migrated
7	Traffic sms alert	Working	Application hosted at Traffic HQ	Not Valid	Internal	Will continue to exist independently post CAS rollout
8	Traffic Mapping System	Working	Application hosted at Traffic HQ	No	Internal /Public	Cannot be integrated as CAS does not contain any GIS based functionality.
9	Automated Vehicle Information System(AVIS) for Traffic vehicles	Working	Located at Traffic HQ	No	Internal	Cannot be integrated with CAS at this point
10	Traffic Monitoring System	Working	Located at Traffic HQ	No	Internal	An Independent system; cannot be integrated with CAS.
11	Teleform Verifier	Working	Server hosted at Traffic HQ	Yes	Internal	Will be integrated with CAS

12	PC Crash	Working	10,000 records per Annum.	No	Internal	An Independent system cannot be integrated with CAS. The Simulation Application is available at workstations in Accident Research Cell, Traffic HQ
13	Web Portal at HQ		Complaint tracking - 1.17GB data of records Court Cases - 344 MB data of records. Servers located at PHQ	Yes	Internal Complaint Monitoring Can be shared with Public	All these functionalities will now be available in CAS. Data will be migrated
14	http://delhipoliceqa.c.in		2.2 GB data of records	No	Internal	Not Part of CAS application. This application will be independent of CCTNS
15	Transfer Process Application			No	Internal	Not Part of CAS application. This application will be independent of CCTNS
16	PCR 100 number facility		Application hosted in the data center at PHQ	Yes	Public Service	This would be a part of CAS Services
17	C4I	Working	Application hosted in the data center at PHQ		Internal	Not Part of CAS application. This application will be independent of CCTNS
18	Inventory Control System	Working	Application hosted in the data center at PHQ		Internal	Not Part of CAS application. This application will be independent of CCTNS