



CHhattisgarh Infotech Promotion Society (CHiPS)

State Data Centre Building, Near Police Control Room,

Civil Lines, Raipur, Chhattisgarh-492001

Tel.: +91-771-4014158

Email: ceochips@nic.in, Website: www.CHiPS.gov.in

Ref: 25/CEO/CHiPS/SRDH/2016

Proof of Concept - Scope Document Chattisgarh - State Resident Data Hub (SRDH)

Date -25th Nov 2016

Table of Contents

1 Introduction	3
2 Background	3
3 State Resident Data Hub (SRDH)	3
4 Objectives of the SRDH	4
5 Scope of the Proof of Concept	4
6 Key Roles and Responsibility of the Stakeholders	6
7 Expected outcomes of the POC	6
8 Process of Carrying out the PoC.....	7
9 Evaluation Mechanism	8

1 Introduction

This document specifies the Background, Objectives and Scope of the Proof of Concept (PoC) exercise with intent to assess the tools and techniques which may support in establishing State Resident Data Hub (SRDH) and further help Government of Chhattisgarh integrate with various departmental databases to SRDH. The organizations participating in the PoC need to demonstrate their understanding and capabilities towards the items listed under Scope of the Proof of Concept. The organization may also showcase any such tool or solution which shall enhance the overall functionality of SRDH and can reduce the turnaround time for implementation of SRDH.

2 Background

CHIPS plans to establish State Residents Data Hub (SRDH) which shall be an authentic, de-duplicated repository of all the residents of Chhattisgarh. The SRDH shall procure UID KYR and Photograph data from the Central Identities Data Repository (CIDR). CHIPS shall also source Departmental databases including domain identifiers for integrating with SRDH. The SRDH shall act as framework for enabling efficient service delivery, better planning and shall also be helpful for monitoring of schemes by Government of Chhattisgarh. The SRDH shall also support in informed decision making based upon the insight derived from the Business Intelligence (BI) and analytics tools proposed to be implemented as part of SRDH.

3 State Resident Data Hub (SRDH)

Unique Identification (UID) Project has been conceived to provide identification to each resident across the country that shall be used as a basis for efficient delivery of welfare and development schemes. Till date, 108 cr. UIDs have been issued. In the state of Chhattisgarh more than 2.6 Crore residents have been issued Aadhaar numbers.

Chhattisgarh plans to rollout Aadhaar based service delivery which shall ensure inclusion of the targeted beneficiaries and help plug the leakages. In view of the above, CHIPS has initiated plans to establish a State Resident Data Hub (SRDH) which shall constitute UID KYR data of the residents of Chhattisgarh. The SRDH application framework shall also be used for seeding UID number in the Departmental database by performing the demographic matching on the UID KYR data and demographic data of the beneficiaries provided by the Department. The UID seeded Departmental database shall enable creation of the enriched SRDH which shall contain UIDAI KYR and the Departmental domain identifier. The enriched SRDH repository shall be further used by the State agencies for better monitoring and planning of various Government schemes. CHIPS envisages that with the implementation of SRDH, Department would not only benefit from easier implementation of the schemes but would also enable them to enhance their monitoring capabilities, ensure targeted benefit disbursement, provide efficient service delivery to the citizens, weed out duplicates, reduce leakages and enhance inclusion of citizen in various welfare schemes.

4 Objectives of the SRDH

SRDH frameworks will provide unique centralized repository containing resident's demographic information with UID Number and scheme identifiers. The repository will facilitate departments to deliver welfare schemes in a much more efficient and effective manner. The key objectives of CGSRDH are:-

- **State Resident Data Hub shall be a Clean, Authenticated and de-duplication data repository for all the Residents of the State** - The Chhattisgarh State Resident Data Hub (CGSRDH) will be the common repository of resident's data that can be used by all State Government departments. The KYR demographic details of the resident shall be sourced from UID database, while the domain identifiers of the Departments shall be sourced from Departmental databases. Collectively the data repository formulated shall be a clean, de-duplicated and authentic master data repository for the State.
- **State Resident Data Hub shall enable efficient service delivery through integration with Aadhaar Authentication** – The CGSRDH apart from creating the master data repository of the State shall also act as a platform which shall be used by the State Government Departments towards seeding of UID number in their databases for enrichment and cleansing of Departmental databases. The CGSRDH shall also provide the platform for Aadhaar Authentication to enable Aadhaar Enabled Service Delivery.
- **Support State Govt. Departments in effective planning of welfare and development Schemes** – CGSRDH shall be a single source of truth which would help individual scheme sponsors, department's personnel and State Planning Department in creating a more robust plans leveraging SRDH. The data repository shall create various reports which would help in planning for Scheme, Department and State.
- **SRDH shall establish the frameworks for effective monitoring of schemes** - Establishing a State level repository of the Residents that would enable better monitoring of various schemes and services provided by the State Departments. Department would have an integrated view towards various benefits being availed by the resident and can ensure that eligibility conditions are met prior to availing the services. Further, the Departments would have traceability of the beneficiaries and would be able to provide targeted services to them.
- **Transformation of service delivery through electronic Know Your Resident (KYR)** – Electronic know your resident service shall transform the way Government services are currently being rendered to the residents. The service shall not only enable Departments to provide efficient service delivery, but shall reduce documentation requirement for service deliver by state government department.

5 Scope of the Proof of Concept

Broadly the scope of the proof of concept exercise shall be consisting of five work streams. These tracks shall enable CHiPS as well as the participating Organization to assess the requirements towards setting up data validation checks, data quality management requirements, standardization of data, integration of databases and creation of KYR plus data and finally running BI/Analytics.

The scope of the proof of concept exercise is to demonstrate the tools and techniques to perform the activities mentioned in the tracks below. It may be noted that the activities mentioned are indicative

in nature and are mentioned to provide guidance the participating organizations in carrying out a structured PoC exercise. The participating organization may customize their approach as per their tool/solution being proposed to be used in the PoC.

1. Validation of data received from multiple data source systems

State Resident Data Hub shall be established as KYR data of the residents along with the Photographs which shall be sourced from UIDAI database while other domain identifiers (e.g. Ration card no., Job Card No., etc.) shall be provided through the State department databases. In reference to the proof of concept exercise, sample UIDAI and State Department data set shall be provided for performing the following activities:-

- Validate data's prescribed format (Essential)
- Validate consistency of data (Essential)
- Detect Syntax errors and highlighting the same (Essential)
- Rejecting files/ records based on rejection criteria for file/ record and System generated auto rejection alarm (Eg. Based in interface Spec, header/ trailer, etc.) (Desirable)
- System generated auto rejection alarm (including reason code, rejection criteria and suggested remedial action) (Desirable)
- Alarm for missing file using checksum etc. (Desirable)

2. Data Quality Management through data cleansing

As a second track of the POC exercise the data needs to be cleansed, duplicate data needs to be removed and incorrect data needs to be highlighted. In the PoC exercise, CHiPS the similar sets of activities shall be required to be undertaken on the dummy data provide to the organization.

- Detection and removal of duplicate data in the database (Essential)
- Detection and highlighting of incorrect information (logical checks on data eg. Age, Date of Birth, name and gender etc.) (Essential)

3. Standardization of data

The third track of the PoC exercise shall enable standardization of the data. This track shall be significantly important as the data from different source systems might have different masters, different formats, different syntax, etc. The outcome of the track shall be to ensure that standardized data shall be loaded in the data warehouse with master as UID which will be integrated at Step 4. Following activities shall be performed as part of the PoC.

- Standardization of Data including Data Types and format (eg. Consistent Mobile number, location, data format etc.) (Essential)
- Incremental/ Differential Data Management (eg. Addition and updation of data from line department) (Essential)

4. Integration of Databases and creation of KYR + data

Once the data repository has been established the Organization needs to showcase various activities which shall enable integration of different Departmental databases by performing the UID seeding of records as follows:-

- 1:1 Matching of demographic data based on common identifiers (Essential)
- Demonstrate most probabilistic matches, 1:N match (Essential)
- Enabling matching of records through photograph matching tools (Desirable)
- UID seeding of records (KYR + data) based on above match (Essential)

5. Business Intelligence Reporting

As a last track of the scope of PoC exercise the participating organization shall be responsible for creating the BI report. Following activities shall be performed as part of this track:

- MIS and Dashboard reports generated from this data repository (Essential)
- Customization of report based on user requirement (Essential)
- Generation of Reports using scheduler (Desirable)
- Auto population of data (based on city name/ village name and pin codes etc.) (Desirable)

6 Key Roles and Responsibility of the Stakeholders

Chhattisgarh infotech promotion society. (CHiPS)

1. Provide access to the dummy data for carrying out the PoC exercise. Actual data scheme at the time of Go-Live may vary.
2. Provide access to CHiPS premises for demonstration of the PoC exercise
3. Provide support and guidance to the participating organizations

Participating Organization

1. Make initial presentation to CHiPS detailing out the approach, tools being used, methodology and expected results.
2. Sign the Non Disclosure Agreement and ensure confidentiality, privacy and security of data/information
3. Carry out the PoC exercise at no cost, no commitment basis
4. Configure, deploy and demonstrate the tools for carrying out the PoC exercise.
5. Seek CHiPS's guidance and support to resolve difficulties faced in the PoC exercise.

7 Expected outcomes of the POC

The POC shall enable the CHiPS to assess the capabilities of the tools and techniques used by the participating organization towards establishment of SRDH, integration of databases and creating report/dashboard as per the business objective envisaged by CHiPS.

The organizations are expected to demonstrate their capabilities with respect to the various activities that are already listed as a part of the scope.

1. Validation of data: The dummy data shall be shared. The participating organizations may also demonstrate capabilities of handling exceptions, if any.

2. Data Quality Management: All logical errors that might be a part of the shared data have to be corrected/ highlighted.
3. Standardization of data: Since the data is expected to be from various source systems, the data may have to be standardized. Beyond this stage the format of the data has to be consistent throughout the application irrespective of the source of data. Also, the participating organizations need to demonstrate how they shall deal with the incremental data that shall come later once SRDH shall Go-Live.
4. Integration of Data – This is the most critical component for this PoC. The evaluation committee shall be interested on observing
 - a. 1:1 exact match of the record i.e. one record of dummy UID data matches with one record of dummy line department data
 - b. Most Probabilistic Match i.e. based on near matches, the tools may highlight the five most probabilistic match
 - c. Dummy UID seeding with KYR + departmental data.
 - d. Finding Golden Records across Homogeneous and Heterogeneous database then building a master database called ‘SERVAM’.
5. The participating organizations need to demonstrate their capabilities with respect to BI and Analytics on the dummy data that shall be supplied.
- 6.OCR Reading for PDF of PoC data.

8 Process of Carrying out the PoC

POC is part of tender document marks, is allotted for Selection of bidder(Ref : SRDH RFP page no 41).The participating organizations are expected to carry out the Proof of Concept exercise on No Cost No Commitment Basis. The cost incurred towards carrying out the PoC exercise shall be borne by the participating organizations themselves. CHiPS does not guarantee any commitment for futuristic actions based upon successful execution of the PoC.

The organizations would be provided with dummy data sets from UIDAI and also a few line departments. The actual data metadata might differ in the live environment. The organizations would have sign non disclosure agreements so as to maintain the data confidentiality. However, the organizations shall be required to source other requirements, if any such as (own hardware, software, licenses etc.). The organizations would have to live demonstrate the tools and the techniques in front of the Evaluation Committee.

During the Initial presentation by the interested organizations, the organizations are expected to showcase their understanding of the project, approach for the SRDH solution, experience and expertise on activity of similar nature including BI and Analytics. Each organization is expected to confirm their participation with the Project In charge at abhijit.c@semt.gov.in. The organizations need to send back the scanned copy of the Cover Letter and Organization Detail. The originals can be sent by courier/ post/ handed over in person on the day of initial presentation. Each organization shall be allotted total 45 mins time for the initial presentation and a round of question and answers, wherein they can clear their issues, if any.

9.Evaluation Technique :CHiPS shall constitute an Evaluation Committee, which shall carry out the entire evaluation process for this POC. The sole purpose of this exercise is to assess the tools and techniques which can support right from establishing State Resident Data Hub to operationalising and maintaining SRDH along with KYR plus data. The committee shall look at the tools and the techniques that are at the disposal of the participating organizations and the effectiveness of the same. The organizations shall be responsible for live demonstration of the scope of the PoC arranging their own infrastructure such as hardware, software, licenses etc. and the committee shall observe the capabilities of the organizations on their systems. All the participating organizations shall be allotted a fixed time slot and the organizations shall be expected to complete their live demonstrations within this time. The organizations will be hard stopped upon the expiry of the time. The organizations are expected to use their allotted time judiciously for this purpose. After the live demonstration, there shall be a round of question and answers.