International Journal of Computational Intelligence and Informatics, Vol. 1 : No. 2, July - September 2011

Integrating E-Governance Systems Using MDM Framework

M. Sundaresan
Department of Information Technology
BharathiarUniversity
Coimbatore – 641 046, India
bu.sundaresan@gmail.com

C. M. Bhuvani
Department of Information Technology
BharathiarUniversity
Coimbatore - 641 046, India
bhuvani.mahalingam@gmail.com

Abstract

E-Governance is an electronic and efficient way of delivering services and information to the citizens. Accurate and up-to-date data exchange gives a tremendous boost to Government to remain agile and competitive. However, the emerging needs of the cross functional data exchange have made data management a complex process, and new generation strategic initiatives like Master Data Management (MDM) helps to establish the standards, policies, framework and protection to organization. MDM is used to manage complex data management scenarios, develop and protect data as an enterprise asset. It also streamline data sharing across systems and provide everyone in the system with a single, consistent view of critical data by using both technology and data governance techniques. This research work deals with redundant E-Governance database is corrected by data profiling, cleansing and compared with existing redundant database. The corrected data is stored in MDM central repository by defining data rules, standards and applied best-of- breed approach in MDM central repository.

Keywords: Master Data Management, Enterprise Management, System of Records, E-Governance.

1. Introduction

E-Governance is a new way of delivering services to citizens. Its main focus is on improving administration efficiency. The increased societal expectations, trust on communication and civic participation have been contributing to the development of new welfare state with knowledge and information access to public [1, 2, 3]. E-Governance systems keep on growing with billions of citizen centric records which is a very big challenge for Integration. There are plans in place to integrate and share the data across these E-Governance systems, which span across different department. These systems are collection of heterogeneous information systems where data is not reliable and inconsistent. This inconsistency of data is leading to incorrect decisions and resulting in more cost to the organizations. The major challenge being faced is, to integrate data coming in from various sources and to maintain data quality in terms of reliability, consistency and accuracy.

Master Data Management (MDM), in its broadest definition, is a technology-enabled process that provides a single context for information. The objective is to create a "single version of the truth" for data collected from disparate systems across the enterprise. It is a central prerequisite for exchanging master data across system boundaries. It is the discipline in charge of creating and maintaining consistent and accurate lists of master data. In many cases, MDM solves complex processes like cleaning and maintaining master data [4].

The paper is structured as follows. In section 2 Existing systems are discussed. Section 3 this proposed systems. In section 4 discusses methodology is presented. Section 5 deals with the conclusions Section 6 deals with the Feature enhancement.

2. Existing System

The existing E-Governance systems have their own implementation and perform designated functions of their own. These systems are up to date in their own context and consolidated information is confusing. These systems are maintaining data as per individual system owner's standards. It is challenging, as and when there is need to integrate or sharing the data across systems. A single view of common data elements, used to integrate E-Governance systems is not available due to inconsistent data standards and resulting in data redundancy [5]. Integrating different E-Governance systems available as of the day will result in the architecture shown in Figure 1.

One system might have updated phone number, other system might have updated address of the person, other system could have updated property details and other system could have updated educational qualifications. But no single system is able to provide all the updated and most recent information of a citizen. This is the point where the need of MDM came up to maintain master data elements in a centralized repository. The increasing complexity of data management environments and increasing demand for cross-function data exchange has created strong need for the MDM programs. The success of an organization is closely dependent on quality, consistent data and the mechanism to share the right information at the right

time. Integrating the master data elements across the E-Governance projects provide single version of truth [6, 7].

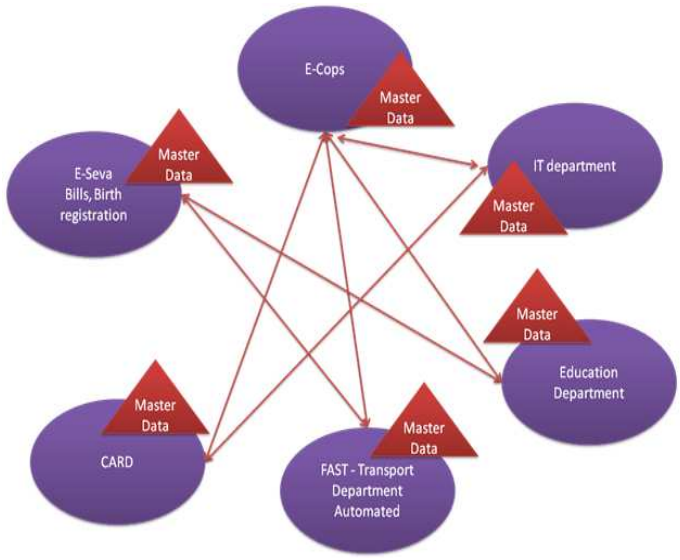

Fig. 1 Architecture of existing E-Governance systems

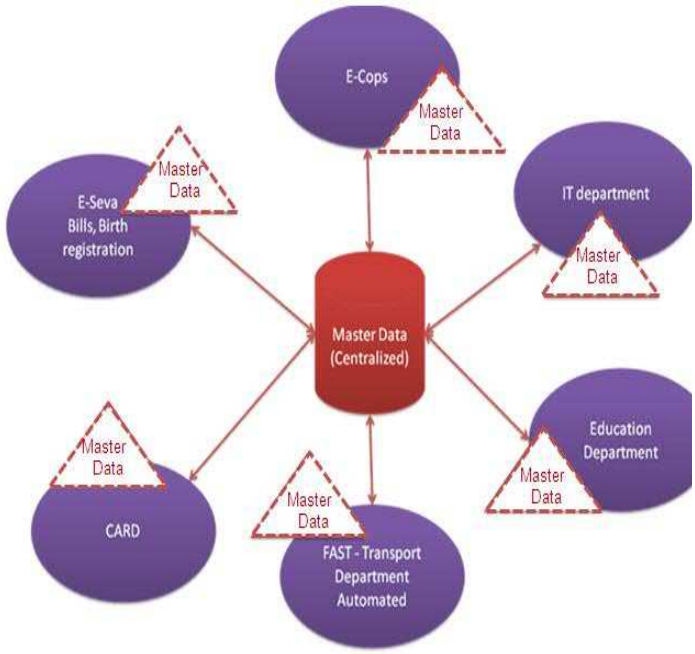

Fig.2 Centralized Master Data System Architecture

3. Proposed System

Data sharing and communication between existing E-Governance systems is not centralized [8]. In case of E-Governance systems, all the information is citizen centric data. Each system has its own version of citizen information. Each system says the data being maintained by their system is true. When looking from the top, the consolidated view is not available and hence resulting in in-consistency, redundancy, duplicity and un-trustable data.

There exists the need, where this data requires be capturing and maintaining in a centralized repository, which needs to be administrated and updated properly. Architecture of the existing E-Governance System needs to be changed as shown in Figure 2, to maintain centralized master data and share the master data across E-Governance systems, hence to maintain single view and version of truth. In case of E-Governance systems data related to citizen will be Master record.

4. Methodology

4.1 Data Profiling

The first step in any MDM implementation is to profile the data. It provides the ability to understand data, highlighting key areas of data discrepancy; to analyze the business impact of these problems and learn from historical analysis; and to define business rules directly from the data [9].

4.2 Data Cleansing

Data cleansing involves standardization, error correction, matching, de-duplication, and augmentation of the data. This process detects the incomplete, incorrect, inaccurate, irrelevant parts of data based on data rule and then replacing, modifying or removing the bad data. The values rejected by a data rule are moved to error table where the cleansing strategies are applied.

4.3 Data Rules Definition

- Remove the white spaces, control characters (enter and back space) from first name and last name
- If first name concatenate with the last name is the First name for another record, then both are same records.
- If previous record does consist normal value and the latest record is empty then copy all previous record values to latest record
- (Eg) Old Record: First Name with enter character – Phone no is available for this record.
- Latest Record---Phone no is not available
- In the above case, copy the phone no from old record to latest record
- If any of two records have the same first name, last name, dob then they are same records
- If any of two records have the same First Name, Last Name And Email Id then they are Same Records
- If Date of Birth (DOB) differs for two People with the same email id always consider first dob entered in e-gov systems.

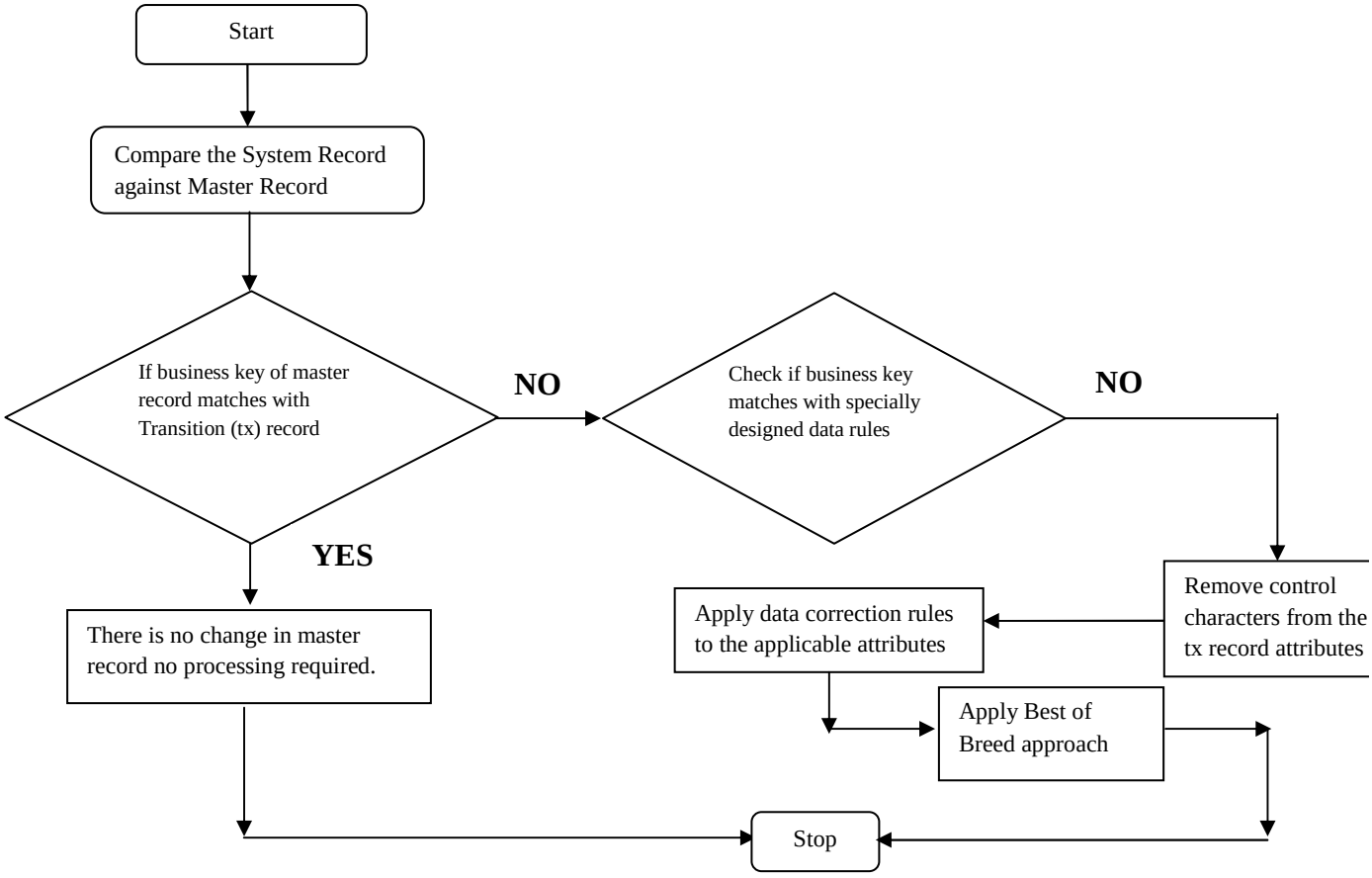

Fig. 3 Flowchart

- If e-seva and e-cops have Two DOB for same person then consider e-seva DOB to update Master record.
- If all source systems contain different phone numbers for the same person, then latest phone number will be considered to update from Master repository.
- If e-cops and FAST have two DOB for same person then FAST has to be considered to update DOB in Master repository.
- If e-seva and e-cops have two different addresses for same person then address from e-seva is taken as to update Master repository, if the modified date of eseva is greater than the ecops.

4.4 Data Integration

Data integration process involves populating the hub database with data from the source systems initially and keeping the source systems synchronized with the hub database as the source systems make changes to the data [10].

Check for duplicates- This process is the heart of most MDM systems. It is both the hardest and most important part of populating the MDM hub. As single view of customer data is required, records describing the same business entity must be combined into a unique record for each unique entity; Duplicate checking process will check for things like alternate spellings and missing words—for example, Raja Rajan, Mr. Raja Rajan, Rajan Raja so forth.

Load the MDM hub database- MDM hub has been loaded by following the steps shown in the Figure 3.If the new record is not already in the hub database then insert the data into the correct tables. But if it is a duplicate, the load process must check the business rules for this entity to decide, what data to update with the incoming record. This process has been detailed in Figure 3 as an example, if there is no address in the current record and the incoming record includes an address, the address is added. If there is already an address and the incoming record also has one, there must be a rule specified to decide which one to keep or if both should be kept. If the business rules can't resolve the conflict, the incoming record should be put on a queue for manual processing. The key of the record should be added to the database to record the connection from the hub record to the source record. This may be used by queries to find the source record or by the hub to publish hub updates to the source systems.

Update the source systems- If loading a new record changes the hub database, the change may need to be propagated to one or more of the source systems. For example, if a new,

address is added to a customer record, other applications that stored information about that customer may want to use the new address.

The load process works best if the most authoritative and complete data sources are loaded first, so that subsequent loads make relatively few changes to the existing hub data. Primarily, however, it's best to record duplicates and synchronize the application data with the hub data. Loading the most critical databases first also leads to earlier time to value.

4.5 Data Synchronization Process

Synchronization is a process that transfers changed master-data records from the source application that made the change to the MDM hub [11]. This introduces the possibility of conflicting updates and inserts from multiple systems, and it introduces some latency between the time a change is made and when it shows up in the MDM hub database; so the business must understand the limitations of this system. After a change has been detected in the source system, it should be sent to the MDM hub as quickly as possible to reduce the update latency. So the frequency of updating master between hub and source applications should be too close to maintain and make available master records in real time.

4.6 Best-of-Breed Approach

Best-of- Breed is a set of different modules collated from different sources to meet the desire process requirements of an organization [12]. In most systems, the rate of change to a given master-data entity is fairly low, so update conflicts should be pretty rare. To reduce the chances of update conflicts, concept of best of breed approach has been introduced into system. Best of breed approach defines which system needs to be considered to update the master data, when data entities are conflicting with each other. Same customer or person record might have two different data of births in e-seva and FAST. In this case Best of breed approach considers data from e-seva system, as it is the place where birth registration information generated from and also it is the most trustable when it comes to DOB. If driving license information is conflicting between e-seva and FAST, best of breed approach considers data from FAST as it is the reliable source of information for this data entity.

4.7 Results- Applying MDM Framework To E-Governance Systems

The results are recorded by applying the MDM frame work on 1101 sample records of ESEVA, 1405 sample records of ECOPS and 1605 sample records of FAST. These E-Governance systems are citizen centric, where all records related to person or citizen, are considered as master data elements. After applying the MDM framework to the E-Governance systems, the findings related to the different attributes of the E-Governance systems are given below. Also all these records are corrected by applying the MDM framework.

The percentage of de-duplicate records found in each system are shown in Figure4. After applying the MDM technique, these metrics shows that, percentage of de-duplicate records vary from 64 to 83 in the E-Governance systems considered to apply the MDM framework. It also shows that, on an average 27% of duplicate data exists in these systems, which will make data volumes grow large in E-Governance systems, in turn the maintenance of these systems, go high. By applying MDM technique these duplicates are removed successfully, while loading data into centralized repository.

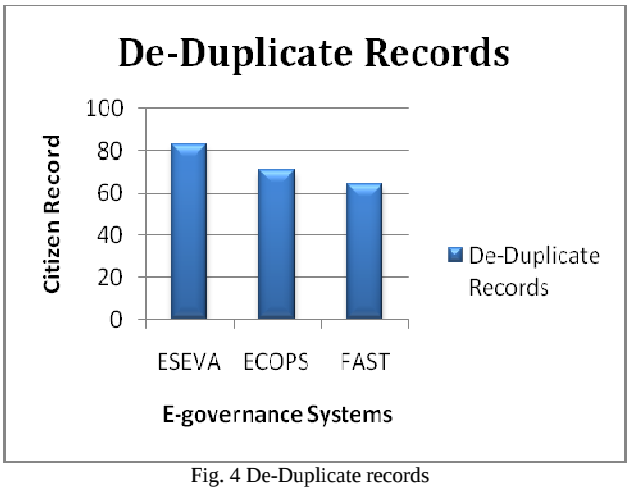

Fig. 4 De-Duplicate records

Results after applying data rules created for E-mail address are shown in Figure5. Data discrepancy for this attribute is 73%, 31%, 56% in ESEVA, ECOPS and FAST respectively. These discrepancies are identified by removing special characters and spaces from the E-mail address and also by applying best-of-breed approach. This kind of inaccurate data might lead organizations in wrong direction. After applying MDM technique and following the steps specified in implementation part, one can able to identify and correct the E-mail address attribute in MDM repository.

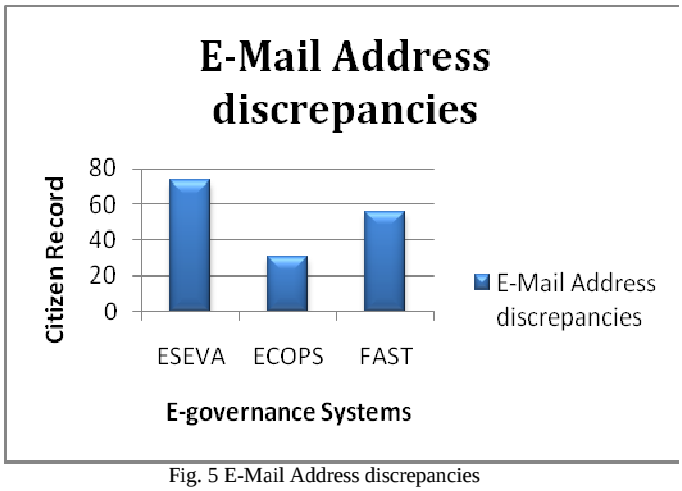

Fig. 5 E-Mail Address discrepancies

Percentage of errors related to first name and last name found in each E-Governance system, are recorded in Figure6. After

applying the MDM technique, these metrics show percentage of error records for attribute first name and last name varies from 11 to 29 in the E-Governance systems considered to apply the MDM framework. These discrepancies are identified by, while applying MDM techniques and removed successfully. Maintaining correct data will always help in taking right decisions and following correct strategies.

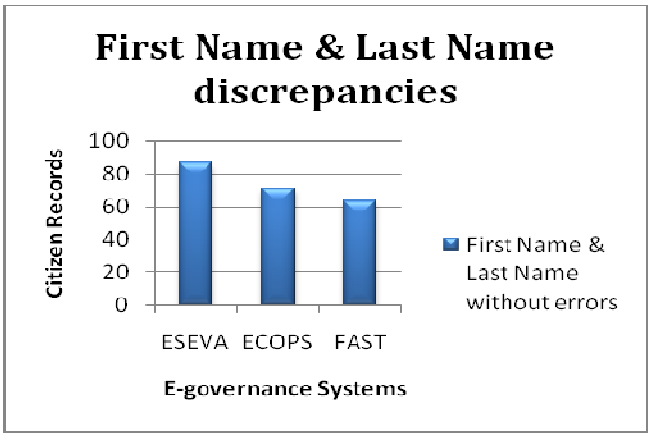

Fig. 6 First and Last Name discrepancies

Results after applying data rules created for phone number are recorded in Figure7. Data inaccuracy for this attribute is 14%, 29%, 21% in ESEVA, ECOPS and FAST respectively. These discrepancies are identified after applying data rules. Also these are corrected while loading into MDM repository. This kind of inaccurate data might lead organizations in wrong direction. By applying MDM technique and following the steps specified in implementation part, one can identify and correct these records.

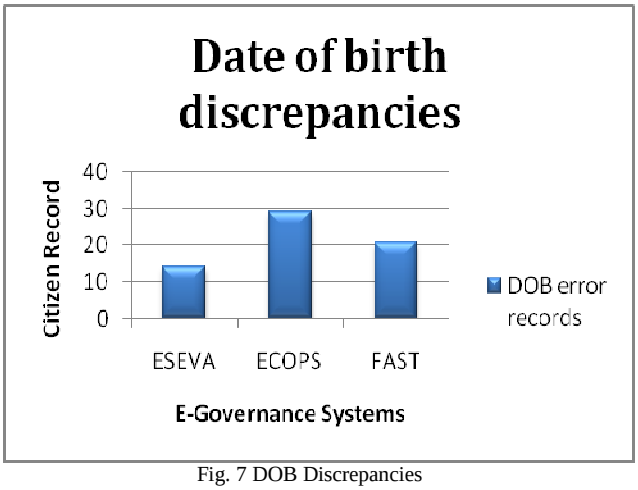

Fig. 7 DOB Discrepancies

Percentage of updates required to apply to phone number attribute in each E-Governance system, are recorded in Figure8. After applying the MDM technique, it is found that, percentage of error records for this attribute varies from 14% to 42% in the E-Governance systems considered to apply the MDM framework. These error corrections are related to either old phone number or phone number as null or new phone number is not updated. Data rules framed for this attribute are applied by using MDM technique and these records are corrected, while loading into MDM repository. Maintaining up-to date data always help organizations to adopt the right choice.

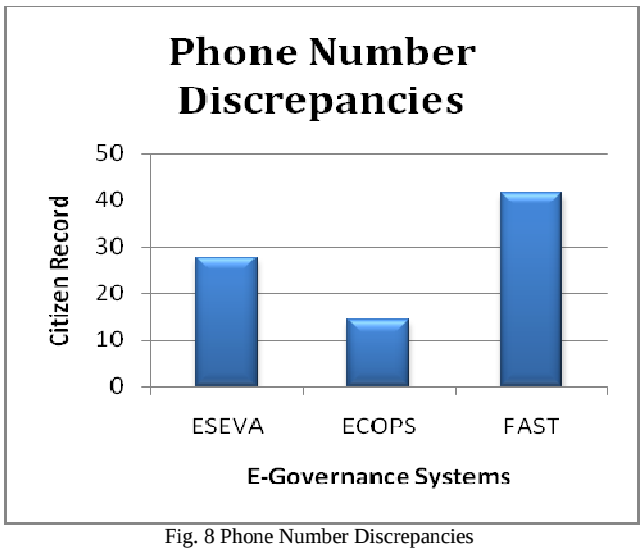

Fig. 8 Phone Number Discrepancies

Below Figure 9 represent the metrics taken by considering all the three E-Governance systems.

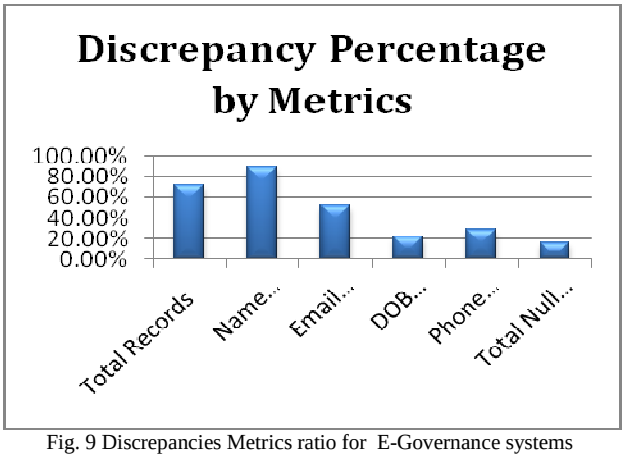

Fig. 9 Discrepancies Metrics ratio for E-Governance systems

As discussed above, all these discrepancies and inaccuracies are removed successfully, while loading master data elements in to centralized repository.

By following MDM framework, it is possible to correct the error records, delete the duplicates and also to keep the central repository up-to date, with single version of truth of the identified attributes. As the master elements are kept in centralized repository, it becomes a painless effort to integrate the existing E-Governance systems.

5. Conclusion

In this research work it has been shown that how the MDM framework helps in correcting the error records, deleting inaccuracies, making the data consistent and available to the application in timely manner. By applying this technique to the E-Governance systems, it is possible to provide "single version of the truth" for the data collected from various E-Governance systems. Now it becomes easy to share data internally and externally, as the information is consistent, accurate and reliable. Instead of maintaining the same data at different places, putting in a centralized repository will save on the infrastructure cost and also the maintenance cost. It is observed from the results that avoiding the duplicates will save lot of space, maintenance and in turn, it saves money to the organizations. In this way MDM framework is incorporated into E-Governance systems successfully and also it is observed that these systems can be managed effectively and efficiently by following this approach**.**

6. Future Enhancement

In this work all the data fields have been used to compare the records in order to identify the errors and discrepancies. All E-Governance systems are citizen centric system. Images or pictures of the users are also stored in these systems. A system needs to be implemented, where in images of different records across different systems need to be compared to identify errors. This will be very much helpful to the MDM approach to identify same records.

MDM framework is applied for ESEVA, FAST and ECOPS. Likewise in future one can do the same to other E-Governance systems like CARD, BHOOMI and Education Departments.

References

[1] Raghunath Mahapatra, P.Sinnakkrishnan, "Internal Management Challenges for Effective E-Governance-An Inidan Perspective", Power Grid Corporation of India., pp.no 176-178, 2004.

[2] Singla, M. L., "E-Governance", Journal of Management Research, vol.no 2,pp.no 165-175, 2002.

[3] Avneet Saxena, S.Wadhwa, "E-Governance and Re-Engineering Issues", Journal of Information Technology, vol.no 13, pp.no 89-95, 2004.

[4] An Oracle Whitepaper, Master Data Management. Sep-2011.

[5] Bhuvani C.M, Sundaresan. M, "Managing E-Governance Systems effectively and efficiently - An MDM framework", International Conference on Emerging Trends in Computing, pp.no 91-95, 2011.

[6] K.B.C Saxena, "Re-Engeering public administration in developing countries",long range planning, vol. no.5,pp.no 703-711, 1966.

[7] M.Hammer, S.Stanton, The Re-Engineering Revolution: a Handbook, Harper business, New York, 2005.

[8] Hajed Al-Rashidi, "Examining Internal Challenges to E-Government Implementation from System Users Perspective",European and Mediterranean Conference on Information Systems ,vol.no 1,pp.no 1-8, 2010.

[9] Amar Jeet Singh, Rajesh Chauhan, "E-Government Databases: A Retrospective Study", Indian Journal of Computer Science and Engineering, vol.no.2,pp.no.66-73, 2010.

[10] Al Merno, Gerg Mancuso, "Master Data Management: Transcending Technology Solving Business Problems", The Journal of Information and Technology Management, vol.no 2,pp.no 9-14, 2007.

[11] Cornel, Christine, "Master Data Management For Collaborative service Processes", Journal of Computer Science, vol.no 1,pp.no 1-6, 2003.

[12] Scott W. Ambler, "Agile Strategies for master data management", The Journal of Information and Technology Management,vol.no 2, pp.no 18-22, Sep- 2007.

Dr. M. Sundaresan did his B.Sc(Applied sciences) Degree at Madurai Kamaraj University. MCA at BharathidasanUniversity, M.Phil and Ph.D at BharathiarUniversity. He has published in Journals also. He is Senior Life member of Professional bodies like CSI, ISCA and ISTE. His areas of interest are Image Processing and Data Compression.

C.M.Bhuvani is currently Research Scholar in Bharathiar University. She did her B.Sc(Computer Science) Degree from BharathiarUniversity, MCA from Tamil Nadu Open University. Her area of interset is Master Data Management. She is a member of ISCA.