# E-GOVERNMENT DATABASES: A RETROSPECTIVE STUDY

**AMAR JEET SINGH**
*Associate Professor*
*Department of Computer Science*
*Himachal Pradesh University, Summer Hill, Shimla-171005 (India)*
aj_singh_69@yahoo.co.uk

**RAJESH CHAUHAN**
*University Institute of Information Technology*
*Himachal Pradesh University, Summer Hill, Shimla-171005 (India)*
rajesh63673@yahoo.com

## Abstract

The government is the biggest producer of information and efficient management of the vast amount of data available within the government departments is of utmost importance. Due to the advent of information and communication technology (ICT), the e-government applications are being implemented to improve government functioning by exploiting ICT potential. In order to meet government data needs efficiently and effectively, the databases need to be designed conforming to standard database design principles. This research paper attempts to study the structure of existing databases used in the government departments for e-government applications. The study reveals the extent of adherence to the most basic parameters while designing the database keeping in view the long term sustainability of e-government applications.

***Keywords:*** E-Government, Database, Information Systems, Relational Table, Database Design Parameters

## 1. Introduction

The government of any country is the biggest producer of data and information (Hood, 1983). In fact the government's business is mainly consisted of data processing and using information within its own departments as well as disseminating it in public for the benefit of the citizens. Thus the job of any government is highly data intensive. Putting in a larger context, the information systems within the government need to be effective and efficient. Information system whether manual or computerized deals with information and its management and due to the advent of information and communication technology, almost all the organizations world over be it business or government are using the technology for faster processing and dissemination of information. An information system is considered as organized combination of people, hardware, software, communications network and data resources that collect, transform and disseminate information in an organization (O' Brien, 2002). Dissemination of information in timely manner, in right format, in right place and with right possessor is needed to make a good information system. Information is the concept unique with the reference to its receiver which means that the same data may convey different meaning to different people unless right format, time, place and authority are not properly agreed upon in advance between the sender and the receiver. Government is a system with hierarchy of subsystems working in tandem for sharing information amongst to achieve multiple goals. Hence there is the need for good information systems within government to disseminate right information amongst the departments, to its constituents, to inform citizens with the policy matters in right mode and to provide government services online. This use of computerized information systems amongst the government departments and reaching out to public for government services delivery is called ICT enabled government or e-government or sometimes interchangeably as e-governance i.e. the use of information system to improve overall governance scenario.

A database is an essential ingredient in information system of e-government with the goal of better delivery of government services by managing the departmental data. Meticulously designed and managed databases contribute to the long term sustainability of the applications in e-governance spectrum by constructing e-government applications which are able to support the various e-government data needs and as well as incremental roll out of e-government in future following the four stages of e-government i.e. internet presence, interaction, transaction and total transformation (Moon , 2002). Since in case of the governments, particularly democratic ones like India, the policies can change frequently in order to meet the citizens' requirements of information, it is imperative to conform to the basic parameters of designing of databases so that scalable, consistent and quality database evolves over the time. The present paper makes an investigation of the structures of existing databases developed over a period of time within government departments to discover as to what extent existing databases adhere to the most basic parameters essential for a good database design.

## 2. Rational of the Study

E-government applications are the part of larger e-governance plan. Almost all governments including that of India have finalized the national e-government strategies and are busy in implementing the priority programmes as seriously as they can within the limitations of administrative and technical shortcomings. It is a moot question

and need to be investigated whether the execution of these applications is going on *ad hoc* basis or the deployment addresses the specific needs of government agencies paying proper attention to the overall need of interaction among the diverse ICT systems within different government departments in order to share and exchange the data. Policy makers are faced with problem of overlapped and uncoordinated data sources. Some inherent complexities are involved in the government data repositories, for the governments hold large amounts of heterogeneous data (National Research Council, 2002) from a wide variety of sources with many different schemas. Hence careful designing is very important in e-government databases. Consistent and meticulously designed database which caters to the future need of e-governance plans can become a great source of good governance including efficiency and revenue increase to government exchequer. Most of the data already lies down in government department as raw data and it needs to be converted into information before government think of making any use of it. In the whole process standard text-book style database design plays important role in overall reliability, scalability and reach of e-government applications.

***Table-1**: Parameters and their effect on database*

| Parameter | Implications |
|---|---|
| 1. Primary key | Consistency, reliability, long term sustainability, robustness, development time, data quality, Integrity. |
| 2. Foreign key | Consistency, reliability, robustness, error rate, Integrity, entity relationship. |
| 3. Data redundancy | Data quality, reliability, increases in the volume of undesirable data. |
| 4. Documentation | Data administration (updating the conceptual schema, usefulness of tables), data migration, data merging. |
| 5. Constraints | Data validation, semantic meaning to data, data duplication. |
| 6. Transactions Handling | Data recovery, reliability, recovery, concurrency control |
| 7. E-R Design | Entity scope, relationship and constraint, Integrity, security, application timeline, |
| 8. Master Data | GUI Development (by providing static data to GUI components for auto completion or validation purpose), Sharing, Interoperability. |

## 3. Methodology

The present study adopts retrospective method of research. Since database design and implementation are the cornerstones of any data intensive application, the authors have collected fourteen databases, consisting of database tables, of e-government applications implemented within last three years by the National Informatics Centre and Department of Information Technology of Himachal Pradesh State/province of India. All databases are application specific i.e. the databases are being used for e-government. These databases are being maintained and managed in proprietary MS-SQL Server. The MS-SQL Server is based upon Relational Database Management (RDBMS) principles. Departments are chosen on the basis of their usability in e-governance scenario. The departments that are chosen to analyse databases are: Transport, Tourism, Legal Services, Irrigation & Public Health, Public Works, Forest, Labour &, Employment, *Panchayati Raj*, *Aanganwadies*, Finance, Education, Economics & Statistics, General Administration and Industry. Consistency, data sharing and reliability are major factors while designing databases because of large amounts of heterogeneous data come from a wide variety of sources (National Research Council, 2002). At the most basic level robustness and proliferation of e-government services depends on the core designing principles of databases and for this reason adherence to the identified parameters viz. Primary, Foreign key, Data Redundancy, Documentation, Stored procedures, Constraints, Transactions Handling, E-R Design and Master-data management are important and are necessary for building fairly good reliable and consistent database for long term sustainability of solution. The parameters and their implications on database design are depicted in *Table -1*. These parameters, though not exhaustive, can be found in any standard RDBMS text book as candidates of a good database design and if adhered to while designing database, these can increase the overall effectiveness of database (Davidson, 2007). The analyses of the relational tables of the databases against these parameters reveal many interesting things for database design and parameter-wise bar graph is presented for all discrepancies. Since in present study the main aim is to know, in totality, about the discrepancies or mistakes against good database design parameters as such

adopted by various departments rather than identifying and co-relating these discrepancies within individual departmental database, the departments are named from A-N. Discrepancies in tables of departmental databases are presented along y-axis as percentage and departments are presented along x-axis.

## 4. Result and Discussion

The subsequent sections discuss the results of the analyses of the chosen databases and critically examine these results.

### *4.1. Primary & Foreign Key*

Databases depend upon keys to store, sort and compare records. The primary key of a relational table uniquely identifies each record in the table. The selection of a primary key is one of the most critical decisions during design of a database table. The other type of key is the foreign key. These keys are used to create relationships between tables. Natural relationships exist between tables in most departmental databases in e-government databases. Defining foreign key helps in making the database consistent. Integrity rule-1 (Entity Integrity) and integrity rule-2 (referential integrity) can only be enforced after defining primary and foreign keys.

The graph in fig 4.1 shows how keys (Primary & Foreign) are maintained in the departmental databases that provide e-government applications. It depicts that large number of databases are lacking in using Primary and foreign keys. This whole scenario leads to integrity problem (Entity Integrity and Referential Integrity) in the database and hence the consistency is compromised which is implemented by foreign keys. Referential integrity ensures consistency during modification in database and prevents users or applications from entering inconsistent data. It means a lot of benefits that can be availed from referential integrity like improved data quality, faster development, fewer bugs and consistency across applications are un-harnessed. Missing the basic requirement of defining the keys compromise the reliability of the services that government is intended to provide through government portal. In practice, this malpractice already has been observed (Blaha, 2008) and lead to buggy software, slow performance and questioning the sustainability of applications.

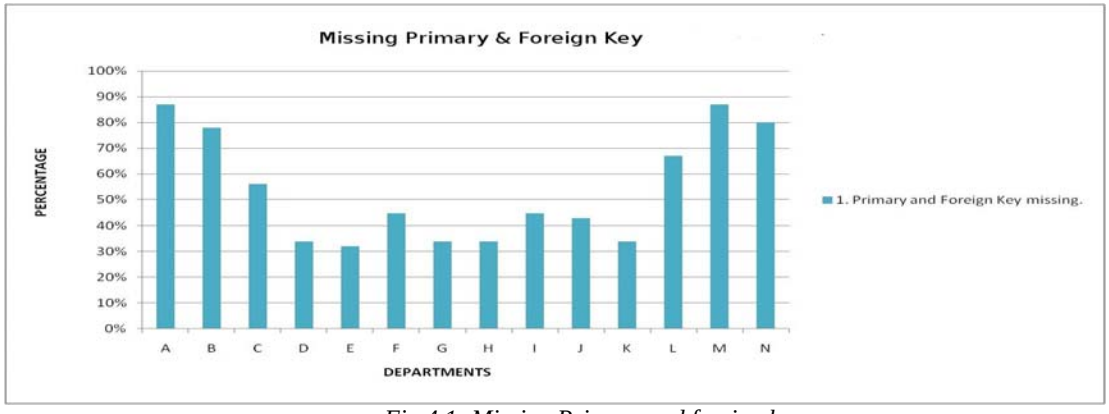


*Fig 4.1: Missing Primary and foreign keys*

### *4.2. Data Redundancies (Normalization)*

The normalization of database tables is a fundamentally requirement of any RDBMS by which the tables are simplified to the extent that attributes are grouped in the simplest way so that changes in individual tables can be made with the minimum impact on the whole database. In other words, redundancies in relational tables are minimized. Redundant data in tables leads to data anomalies, particularly modification anomalies. Normalization helps in removing redundancies. Properly normalized data not only reduces duplications but also helps in removing deletion, updation and insertion anomalies. It is a trade-off that up to what level normalization is performed to avoid anomalies (Tangwongsan et. al., 2010). This is because after the normalization, total number of tables increases and retrieval of data takes more time as the join operation is time wise costly operation. Generally the Boyce Code Normal form is fastest way of normalization database to minimize redundancies, however as per the given databases analysis as shown in fig 4.2, almost all databases are redundancy infected and six departments show significantly high redundancies (more than 50%) which are a concern for reliability of e-government database.

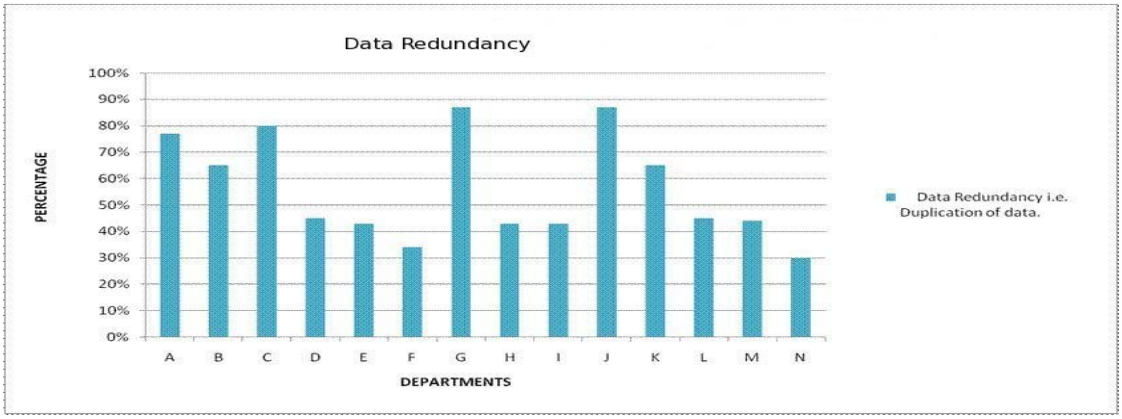


*Fig 4.2: Data Redundancy*

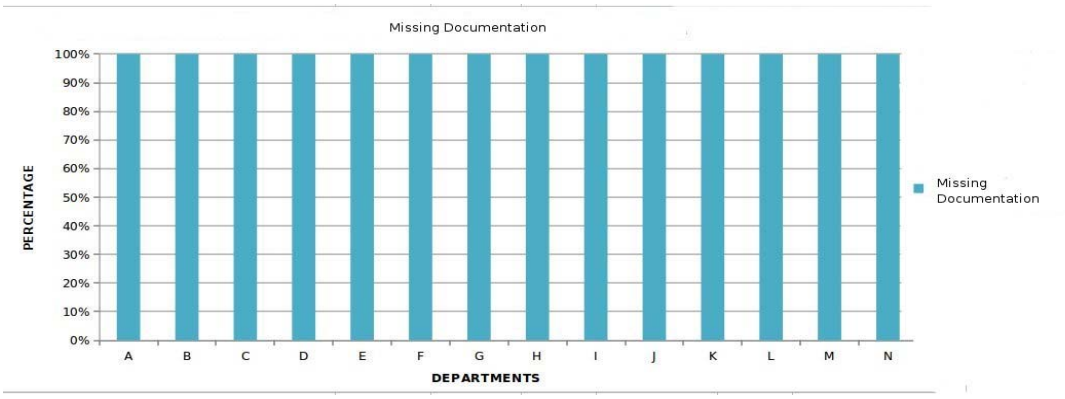


*Fig 4.3: Missing Documentation*

### *4.3. Missing Documentation*

Documentation is one of the most important parameters in any software particularly in databases which is almost analogous to the blue print of a building design, which helps in conveying the meaning of data to the management and other concerned technical personnel. Documentation of database helps in understanding the logic behind database design in many real life scenarios viz. for future modification of database, migrating the database, merger and acquisition of organizations, updating the conceptual schema or absence of original database designer during future scaling up of database. In regards to relational databases adhering to naming standard is a property of well-designed data model, definitions of tables, columns, relationships and even default and check constraints it contains are clearly need to be mentioned so that the intention of its use is clear to everyone. However as shown in fig 4.3 the analysis reveals that all the databases have no documentation either in the database or in any other form. The database designer concerned with the application is the sole individual who knows the logic behind the database design and s/he also relies on his brain only. No format for documentation is followed at departmental level. Although documentation does not directly affect the quality of service but it mars the prospect of future database modification as in government set-up requirements of modification to existing information needs change quite frequently. The poor documentation also hints that designing a database for a government department is 'single-man business' lacking any team work and software engineering approach.

### *4.4. Constraint Enforcement*

A constraint in database design enforces the rule for the values that could be entered or rejected in a database table while insertion, updation operations are performed. It is a good practice to validate data before entering it to the database, especially when multiple applications/users access the databases. Databases are checked for

three constraints 1) Null Constraint, 2) Unique Constraint, 3) Check Constraint.

- Null constraint adds semantic meaning to data model. It provides a way to define that column must have some value. Neither insert nor update operation can be performed without adding a value to this field. This also removes ambiguity e.g. *"find all employees who are not officer"*. Have query "select ***** *from emp where job <> 'OFFICER' "* which is wrong, correct one is "select * *from emp where job <> 'OFFICER' or job is NULL;"*
- The unique constraint doesn't allow duplicate values in a column. If the unique constraint encompasses two or more columns, no two equal combinations are allowed. However, if a column is not explicitly defined as not null; nulls can be inserted multiple times.
- A check constraint allows stating a minimum requirement for the value in a column. It is also possible to state a check constraint that checks the value of more than one column.

Fig 4.4 shows that very less number of tables in e-government databases use constraints, which means that data can not be validated at the time of data entry. If the data is validated at the source, error rate will be less and data is more reliable.

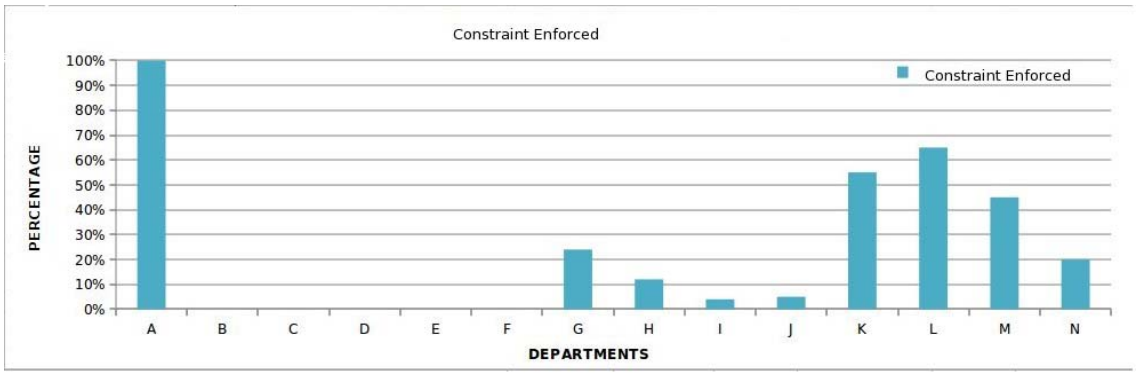


*Fig 4.4: Constraint Enforced*

### *4.5. Transactional Inconsistencies*

Transactions as a facility are provided by almost all the proprietary RDBMSs and help in maintaining the consistency of databases in multi-user environment where many read and write requests concurrently execute on a single database. Transactions ensure the consistency, durability and integrity of database. Do, undo and redo operations with ACID (Atomicity, Consistency, Isolation, Durability) properties helps in avoiding inconsistencies in database and database always remains in legal and valid state. Transaction facility also contributes towards concurrency and data recovery. Analysis as per fig 4.5 shows that most of the departments are reluctant in using transaction facility. Bars corresponding to high percentage mean more reluctance in using transactions. Avoidance of transactions by the department means that they have no procedure handy to recover the previous data or state. This also implies that e-government databases may give inconsistent and hence invalid results at anytime during the lifetime of the database.

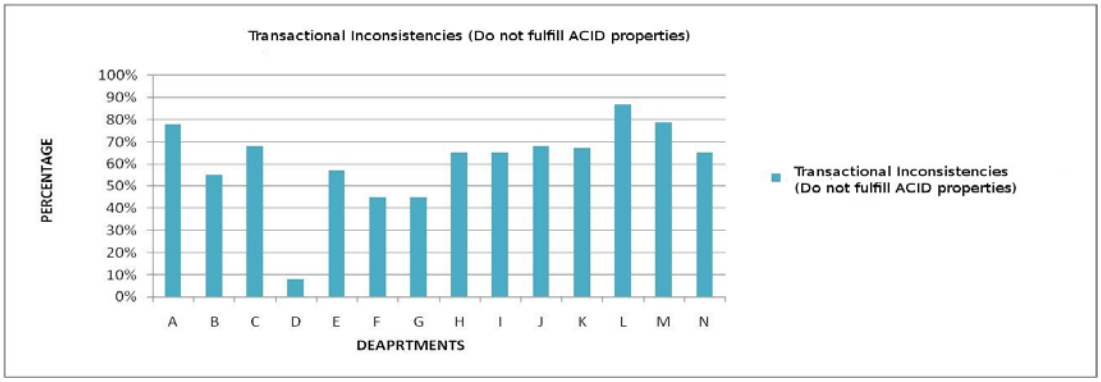


*Fig 4.5: Inconsistencies (ACID Properties)*

### *4.6. E-R Design*

Conceptual designing is a major task in whole database design process and in an e-government initiative it shows capability of rendering services efficiently to citizens in present and future. Conceptual design makes the design more comprehensive, easy to understand and can be used to communicate the intent of database design with non-technical user. It increases expressive power of designer (Batini et al. 1992; Biskup 1995; Hull & King 1987). If the designer cannot visualize the holistic view of the enterprise data, it is very difficult to add new applications in future without modifying current database structures and thus affecting their performance as well as availability. Substantial changes in the database structures could have a huge negative impact on the whole application in terms of money and time loss. On the other hand conceptual design shows the information requirements of user and includes total number of entities involved, their relationship and constraints in the database. Empirical studies suggest that the misrepresentation of relationships is a common error in the design process which affects the reliability of database (Dey et. al., 1999). Entity-Relationship (E-R) diagrams are the simplest tools used to comprehend and model all the entities and their relationships amongst in any database design. Tools and techniques for common problems like many-to-many relationships and multiple-object associations, repeated attribute associations, reflexive or recursive associations etc. are available for the use to build robust and reliable databases (Stephens, 2009). Fig 4.6 shows the conformance to the E-R design in government databases. Out of 14 databases 9 databases have little conformance to conceptual design. This scenario is beckoning to the fact that most of the departments have not developed any conceptual schema and databases are designed on *ad hoc* manner.

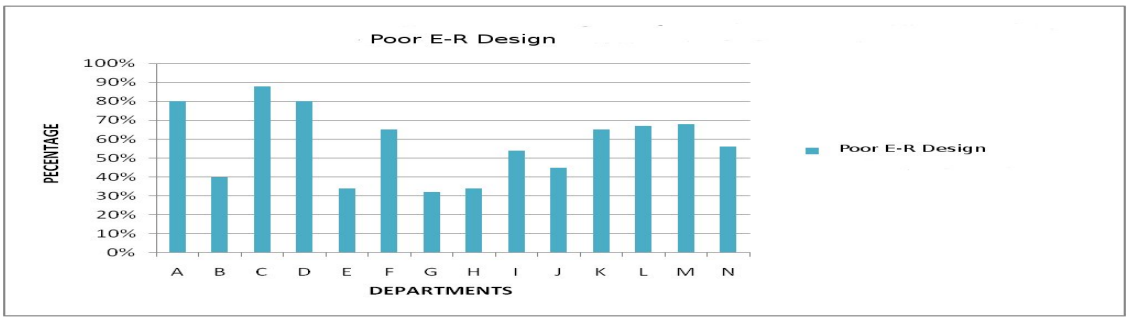


*Fig 4.6 E-R Design*

### *4.7. Master Data*

Master data denotes any organization's essential basic data which remains unchanged over a specific period of time. In practice, however, master data frequently lacks consistency. A general issue in master data management at organizational level is in ensuring data quality in terms of consistency. Obsolete (non-current) and inconsistent master data can lead to error-prone operations (Loser et. al., 2004). Master data repositories also facilitate interoperability of application and sharing of data across the departments or among the applications (e-Government Unit, 2006). An application in development stage also needs master data in Graphic User Interface (GUI) components (combo boxes etc) for auto completion and validation purpose. Master data, if maintained centrally facilitates the sharing of basic data about the organization among the applications and acts as a basic data for processes. The distribution of master data is always initiated by the central system. Inconsistent master data in any information system leads to incorrect decisions, reduced satisfaction and high costs in service (Loser et. al., 2004). Fig 4.7 shows that no master data has been maintained separately by any of the department and the non-availability of master data limits sharing of data across different government departments and scope of services provided by the departments. E-government is comparatively new and growing area and it may be easier to start with few key sources of master data and expand effort once success has been demonstrated and lessons have been learnt. As the scope of e-government extends, the size of database also increases and the need for sharing of databases across different department also arises. In order to access previous transaction records maintaining master data is of immense value in government sector for decision making and legal purposes.

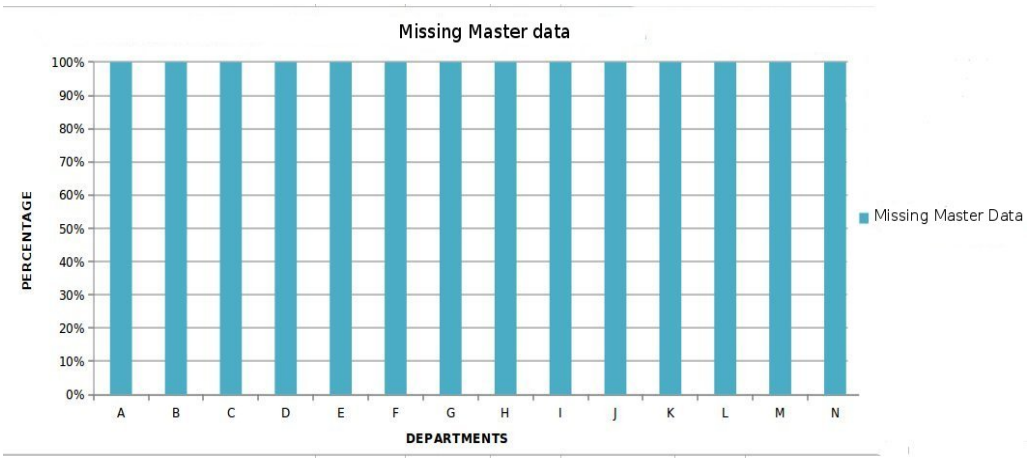


*Fig 4.7: Master Data*

## 5. Conclusion

The analyses of the chosen databases show that the development of e-government applications invariably lacks conforming to standard database design parameters. The study annotates that data in government information systems is being managed on *ad hoc* basis whereas a single technical person is responsible for all aspects of database handling without any documentation, thus the government information systems are not scalable at the moment. The lack of master data and poor design also points out that these databases are not worth for integration and at present e-government applications are just isolated experiments within the government departments. If put to actual use at large scale, the poor database design will lead to various problems like data inconsistency, inefficient performance of data access, partial comprehension of the needs of the e-governance facility users and it will hamper future extensibility and integration of the existing e-governance facilities being envisaged. However the scope of the present probe is limited to a few e-government databases within a State and hence the results of the study are only precursor to further research in a larger context of e-government databases.

.

## Acknowledgment

The authors sincerely express their gratitude to the ICT professionals working in the Government of Himachal Pradesh who provided access to e-government application databases and frankly discussed the various issues about database design within the government departments.

## References

[1] Batini, S.; Ceri S.; Navathe, B. (1992): Conceptual Database Design: An Entity Relationship Approach. Benjamin/Cummings, Redwood City, CA.
[2] Biskup, J.; Menzel, R.; Polle, T. (1995): Database schema design theory: Achievements and challenges. In Proceedings of the 6th International Conference on Information Systems and Management of Data (Bombay, India, Nov.), pp.14 – 44.
[3] Blaha, M. (2008): Referential Integrity is important for databases, Modelsoft Consulting Corp, www.odbms.org/experts.aspx., Accessed on: 15 July 2010.
[4] Davidson, L. (2007): Ten Common Database Design Mistakes, http://www.simple-talk.com/sql/database-administration/ten-common-database-design-mistakes. [On-line], Accessed on : 22 July, 2010
[5] Dey, D.; Storey, V. C.; Barron, T. M. (1999): Improving Database Design through the Analysis of Relationships, ACM Transactions on Database Systems, Vol. 24, No. 4, December 1999, pp. 453–486.
[6] e-Government Unit(2006): e-Government Metadata Standard, Version 3.1 29 August 2006, http://www.govtalk.gov.uk, Accessed on 12 Jan.,2010.
[7] Hood, C. (1983): The Tools of Government, Mcmillan.
[8] Hull, R.; King, R. (1987): Semantic database modeling: Survey, applications, and research issues. ACM Comput. Surv. 19, 3 (Sept. 1987), pp. 201–260.
[9] Loser, C.; Legner, C. & Gizanis, D. (2004): Master Data Management for Collaborative Service Processes, http://www.alexandria.unisg.ch/EXPORT/DL/Christine_Legner/28091.pdf, Accessed on 28 Jan., 2010.
[10] Moon J. M. (2002): The Evolution of E-Government among Municipalities: Rhetoric or Reality? , Public Administration Review, Vol. 62, No. 4 (Jul. - Aug., 2002), pp. 424-433, Published by: Blackwell Publishing on behalf of the American Society for Public Administration.

URL: http://www.jstor.org/stable/3110357, [On-line], Accessed on: 10 March, 2010.
[11] National Research Council (2002): Information Technology Research, Innovation, and E-Government, National Academy Press, Washington. D.C.
[12] O'Brien, J; Marakus, G. (2009): Management Information System, pp. 48, McGrawHill Pvt Ltd.
[13] Stephens, R. (2009), Beginning Database Design Solutions, Wiley Publishing Inc, 10475 Crosspoint Boulevard Indianapolis, IN 46256, pp. 186-206.
[14] Tangwongson, K.; Himabindu, P.; Kaminsky, M. (2010): Efficient Similarity Estimation for Systems Exploiting Data Redundancy. In Proc. IEEE INFOCOM, (San Diego, CA), Mar. 2010.