

**DIGITIZATION OF CENCUS SYSTEM INCREASES SYSTEM EFFECIENCY
AND DATA AVAILABILITY**

**SAKIB MD. ARAFAT
ID: 173-25-629**

This Report Presented in Partial Fulfillment of the Requirements for the
Degree of Masters of Science in Computer Science and Engineering

Supervised By

Dr. Sheak Rashed Haider Noori

Associate Professor

Department of CSE

Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

SEPTEMBER, 2018

APPROVAL

This Project titled “**Use of digital system in census increase efficiency and reduce cost**”, submitted by ‘Sakib Md. Arafat’ to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of M.Sc. in Computer Science and Engineering (MSc) and approved as to its style and contents. The presentation has been held on 19th September, 2018.

BOARD OF EXAMINERS

(Name) [Font-12, Bold]

Chairman

Designation

Department of ETE [Font-12]

Faculty of Science & Information Technology

Daffodil International University

(Name)

Internal Examiner

Designation

Department of ETE

Faculty of Science & Information Technology

Daffodil International University

(Name)

Internal Examiner

Designation

Department of ETE

Faculty of Science & Information Technology

Daffodil International University

(Name)

External Examiner

Designation

Department of -----

University of Dhaka

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Dr. Sheak Rashed Haider Noori, Associate Professor, and Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Dr. Sheak Rashed Haider Noori
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Submitted by:

Sakib Md. Arafat
ID: 173-25-629
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty Allah for His divine blessing makes us possible to complete this project successfully.

We really grateful and wish our profound our indebtedness to **Dr. Sheak Rashed Haider Noori, Associate Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “*Data Science*” to carry out this project. His endless patience ,scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to Prof. Dr. Syed Akhter Hossain, professor and head, Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Census is a process of collecting data, compile and analysis them for publishing demographic, economic and social data at a specified time of a country. Census for a country has various types. Population census is one of them. In Bangladesh population census is run in every ten years. But in ten years population increases in a large number. Along with population various social and financial marker also fluctuates. These markers are very important for a country to take various important decisions. As the current census is a manual black and white process. It is a time consuming process.

The objective of our work is to digitize the process, which will be less time consuming and also reduce the expenses for the process to be completed. So that government can run the census in a short interval of two to three years. These will keep the data markers up to date. Both the government and non-government organization will be benefited by raw data for medical, social or other types of analysis.

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
CHAPTER	
CHAPTER 1: INTRODUCTION	1-3
1.1 Introduction	1
1.2 Motivation	2
1.3 Rationale of the Study	2
1.4 Research Question	2
1.5 Expected Output	3
1.6 Report Layout	3
CHAPTER 2: BACKGROUND STUDY	4-6
2.1 Introduction	4
2.2 Related Works	4
2.3 Scope of the problem	6
2.4 Challenges	6

TABLE OF CONTENTS

CONTENTS	PAGE
CHAPTER 3: RESEARCH METHODOLOGY	7-14
3.1 Introduction	7
3.3 Data Collection Procedure	7
3.4 Statistical Analysis	11
3.5 Implementation Requirements	12
CHAPTER 4: Experimental Result and Discussion	15-18
4.1 Introduction	15
4.2 Experimental Results	15
4.3 Descriptive Analysis of Improve Efficiency	17
4.4 Descriptive Analysis of Data Availability	18
CHAPTER 5: Conclusion	19
5.1 Pros and Cons	19
5.2 Future Plan	19
5.3 Conclusion	19
REFERENCES	20

LIST OF FIGURES

FIGURES	PAGE
Figure 2.1: Paper base process of population census	5
Figure 3.1: Proposed digital census system	8
Figure 3.2: Allocated Menu	8
Figure 3.3: Permission Allocation	9
Figure 3.4: Team Allocation	9
Figure 3.5: Housing Census	10
Figure 3.6: Population Census	10
Figure 3.7: Data Collection Form	11
Figure 3.8: Graph of times to input data of families consisting of different members using paper based system	15
Figure 3.9: Graph of times to input data of families consisting of different members using our designed prototype	16
Figure 3.10: Graph of differences between data input time of regular paper based system and using our prototype of digital census system	17

CHAPTER 1

INTRODUCTION

1.1 Introduction

To know about the social, financial and health condition of the population of a country census is a must. Census is the only process by which data is gathered properly by surveying all over the country under the surveillance of the government. Then the data are processed and analyzed to end up with some important marker.

These data markers are very useful to the people who work with data analysis or statistics to provide various information about the population, progress, earning person, rate of birth and death, average life expectancy and health condition analysis. In Bangladesh population census is run in every ten years ^[1]. Where UN recommends a census enumeration at least once every ten years, and once every 5 for even better data, rather than simply relying on estimates and projections alone ^[2]. According to the condition of the country these markers may fluctuate randomly.

After every ten years period, we only get the data of fluctuations average for ten years. There may be some impacts of any uncertain incidents in a specific year between the ten year periods for which markers may changes rapidly. Our goal is to be more accurate and store data of every single year by digitizing the population census using new technology and make the work easier for the both official and field workers.

1.2 Motivation

Our motivation is to use modern technologies to digitize census process. This will make the process faster and easier. Using modern technologies we can achieve higher accuracy. It will reduce the cost and the time of the full process and data will be stored in a central database. So that related workers do not have to gather data and then store in a central system.

1.3 Rationale of the Study

The population census is run by a group of field workers under surveillance of government. They go door to door to collect information about every family. This is the actual process which is run in almost every country in the world. In the world of digitization this process is still in a black and white process. Less use of technology has made the process slow and very complex. From the official worker to the field workers both have to be very careful to their work, otherwise the whole process has to be performed again. The paper works are hard to manage and they have to carry forms for each family they have to visit on a working day.

1.4 Research Question

We believed that digitizing the census process then it will be less time consuming, cost effective, more efficient and less work force will be needed. To achieve our goal we fixed some research questions. Our research questions were the followings:

1. How a digital system can reduce the time of census system?
2. How it can improve the process efficiency and open the scope of data availability?

1.5 Expected Output

At the end of the study we expected to have a result of a system that ensures a cost effective, high data accuracy and use of fewer workforces.

1.6 Report Layout

Our objective is to develop a prototype for a digital system and show the analysis result difference of regular paper based system and online digital system. The report is designed in a way that in the next chapter we will know about the background studies and some related works about the system we are proposing here. Then we will discuss about the research methodology of our work and after that the chapter will contain the experimental analysis of our system with regular census system.

CHAPTER 2

BACKGROUND STUDY

2.1 Introduction

In this chapter we will discuss about the background of our study. Some related literature along with the scope of problems will be discussed. There will be some challenges in this system. We will discuss how to minimize the unexpected problems and overcome the challenges.

2.2 Background Study

In every country population census is done by the government. Generally national statistical office runs by the government, manage the whole census process. This is a very time consuming process where all the works are started from a central office. The information collection forms, area selection forms, field worker and their managerial information forms are printed and then the related papers, documents of information and instructions for the team managements and also the field workers are sent to the regional sub offices. If there is any fault in the forms or documents, they have to go through the same process again.

These forms and document then distributed among the small teams so that they can start their survey. They have to go door to door carrying forms for each family. After information collection they have to return the survey form to their regional offices. After collection of data if any unwanted incidents occur such as they forms are lost or damaged they have to survey again using new forms.

After collection of data of an area the papers are returned to the main office for the next step. In this step gathered data are processed and analyzed for further use and

publications. In the main office the data is stored often in computer document formats. To analyze them for a particular research is quiet hard.

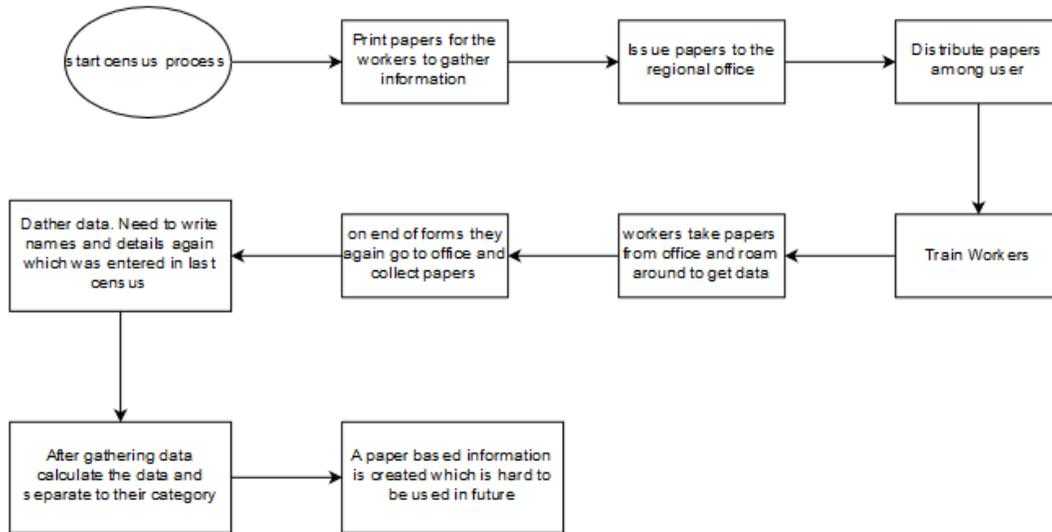


Fig: 2.1 Paper Based Census System

From this study two main issues are found. Paper based census process is slow and less quantity of gathered information. The result of the average of five or ten years can be found here. But big micro data are still inaccessible as they are paper based. There is also two circle of error. If these are reduced from the process, any correction or data damage can be overcome along with time. The process has bidirectional document transportation phases. In some cases multiple times of transportation phase, if there is any error from the central office.

Full data has to reach to the end point for the calculation to be stored. Process delay of any regional office will cause a delay to the whole census process. As the total process is a manual process data calculation may have some human error. Though it is a huge number, so these errors are negligible.

Census Digitization has been accepted in many countries of the world. In the year of 2010 Brazil conduct a digital census ^[3]. In 2017 Swaziland successfully undertakes first digital census ^[4].

2.3 Scope of the Problem

The forms used in the census can gather less information. A full functioning webpage can replace the use of paper. Then transportation of the documents and the papers to the regional offices is not required here. Using simple tablet computer field workers can collect the data. In that case expenses may be raised. But it is reusable for multiple times. If any error occurs the whole nation paper need not to be changed. Here expenses reduce. Data is stored in central database. So, after data collection these data does not need to be sent to the central office for compute and analyze. Complex computer software can continuously process data. In that case everyday data collection progress can be monitored.

2.4 Challenges

It is a new concept and total digital system so that first challenge will be the setup of the whole system according to the present system. There will be hierarchy of administrative system which will have to maintain to get better result. The use of new devices may create some problems initially. Proper training will overcome the challenges.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

In this chapter we will discuss about our research instrument, data collection procedure, statistical analysis of paper based and digital system. At the end of the chapter we will discuss about the whole system implementation and training.

3.2 Development of Prototype

We have developed a prototype for the experiment. The prototype was based on the form and the information of data collection, gathered from the bureau of statistics website and officials. The process the follows to perform a census system was used as the knowledge base for the prototype design.

Digitization of census system will work over the internet so we chose web application to build the prototype. The prototype user interface was designed in HTML using Bootstrap a CSS framework. MySQL was used for data storage. We have used PHP as the backend language to store and fetch data in the web application.

The prototype was divided in to two parts. First one is to build the team for the field workers. That is why we have to create a central administration who performs the creation of regional administration, team leader for the field work force and the members for the team leaders.

The second part was developed for gathering data. Data was gathered in two parts. According to the form of the statistics bureau the form has been formatted into two section. One gets the information about the house and another section gets the information about the person lives in the house.

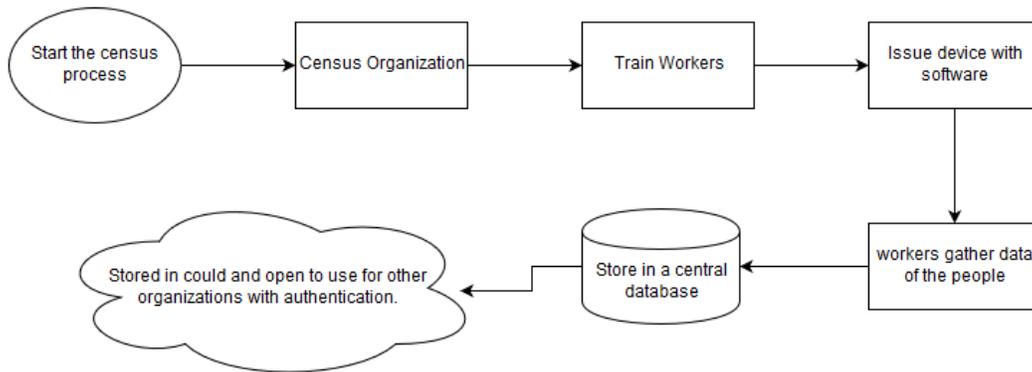


Fig: 3.1 Proposed Digital Census System

The screenshot shows the user interface of the Digital Census system. At the top, there is a navigation bar with links for 'Home', 'About', 'Contact', and 'Census Panel', along with 'User Settings' and 'Logout' options. Below the navigation bar, a header displays user information: 'Location : Dhaka', 'Code : ca00123', 'Name : sakib', and 'Type : Central Admin'. The main content area is titled 'User has the following permission' and lists 11 permissions in three columns:

User has the following permission		
1. Permission	2. Set Division	3. Set District
4. Set Sub-District	5. Set Thana	6. Add Regional Admin
7. Add Team Leader	8. Set Area	9. Add Team Member
10. Team Allocation	11. Start Census	

Fig 3.2: Allocated Menu

Digital Census Home About Contact Census Panel User Settings Logout

Location : Dhaka Code : ca00123 Name : sakib Type : Central Admin

Permission

User Type:

User Code:

User Type:

Fig 3.3: Permission Allocation

Digital Census Home About Contact Census Panel User Settings Logout

Location : Dhaka Code : ca00123 Name : sakib Type : Central Admin

Team Allocation

Division:

District:

Sub subDistrict:

Thana:

Fig 3.4: Team Allocation

Digital Census Home About Contact Census Panel User Settings Logout

Location : Dhaka Code : ca00123 Name : sakib Type : Central Admin

Housing Census

Area:
--Select Area--

Address:
Address

Serial No:
Serial No

Housing Category:
General

Wall Type:
Bamboo

Roof Type:
Bamboo/Polythin

Ownership:
Owner

Water Supply:
Tap

Toilet Facility:
Sanitary

Electricity Facility:
Yes

Cultivable Land:
Yes

Earning Source:
Cultivation Land-own

Is it a tribal house?
Yes

Submit

Fig 3.5: Housing Census Data

Digital Census Home About Contact Census Panel User Settings Logout

Location : Dhaka-Dhaka-Dhamrai-Dhamrai Code : teamMember001 Name : teammem1 Type : Team Member

Population Census

House Code : houseCode00000013

Name:
Sakib

Age:
25

Relation With Head:
Child

Sex:
Male

Marital Status:
Married

Religion:
Islam

Highest Education:
Graduate

Field of Education:
Technical

Does go to school:
Yes

Capable of writing letter?:
Yes

Job Field (Last one month):
Service

Designation:
Employee

Nationality:
Bangladeshi

Living Status:
Yes

Moved to abroad:
No

Migrate:
No

Submit

Fig 3.6: Population Census Data

3.3 Data Collection and Process

We have collected data in two steps. First regular paper based system and second is our proposed digital system. To gather data we have work with two volunteers and data will be taken from ten families consists of different family members.

Two volunteers have taken data from these families separately. First we are discussing the process for regular paper based system.

For the first step, we have collected a form of population census and produced multiple copies. The volunteers have gone to the separate families with the data entry form. First volunteer collected data and second one has captured the time.

The image shows a population census form with a header in Bengali. The form is divided into several sections. The first section contains personal details such as name, gender, date of birth, and marital status. The second section is for household information, including the address and the number of family members. The form is filled out with handwritten data for several individuals. The columns are organized into groups, with some columns for identification and others for demographic data. The form is titled 'Appendix 2.1. Census Questionnaire' and has a page number '342' at the bottom right.

Fig 3.7: Population Census Form

In the second step, the volunteers have gone to the same ten families and gather the data using our developed prototypes. Like previous system one has gather data and another captured time.

3.4 Statistical Analysis

Table 1: Collection of data using paper based system

Family	Family Member (m)	Paper Form	Time (t) In minutes	New Information
Family - 1	5	Volunteer - 1	12.35	No
Family - 2	2	Volunteer - 2	6	No
Family - 3	4	Volunteer - 1	9.56	No
Family - 4	4	Volunteer - 2	10	No
Family - 5	6	Volunteer - 2	20.34	No
Family - 6	4	Volunteer - 1	9.27	Yes
Family - 7	4	Volunteer - 1	9.35	Yes
Family - 8	2	Volunteer - 1	6.5	No
Family - 9	5	Volunteer - 2	11.07	Yes
Family - 10	3	Volunteer - 2	8.5	No

The above table represents the data of ten separate families consists of different members. We gathered data shuffling the volunteers to achieve a proper analysis by paper based method.

Table 2: Collection of data using online prototype system

Family	Family Member (m)	Digital Form	Time (t) In minutes	New Information
Family - 1	5	Volunteer - 2	7	No
Family - 2	2	Volunteer - 1	3.55	No
Family - 3	4	Volunteer - 2	6	No
Family - 4	4	Volunteer - 1	7	No
Family - 5	6	Volunteer - 1	13.20	No
Family - 6	4	Volunteer - 2	5.49	Yes
Family - 7	4	Volunteer - 2	6.4	Yes
Family - 8	2	Volunteer - 2	3.5	No
Family - 9	5	Volunteer - 1	7	Yes
Family - 10	3	Volunteer - 1	6	No

The above table represents the data of ten separate families consists of different members. We gathered data shuffling the volunteers to achieve a proper analysis by our developed prototypes.

From the gathered data we can get the average value for both of the methods. To get the average of a data we can use ‘Arithmetic Mean’ as our statistical analysis method.

To calculate the average time of paper based system, we denoted the arithmetic mean as AM_{ps} . Let the time for each family be t and number of total member of all the families be n ,

$$AM_{ps} = (t_1 + t_2 + t_3 + \dots \dots \dots + t_{10}) / (n_1 + n_2 + n_3 + \dots \dots \dots + n_{10})$$

Putting the values from table one, we get the value of average mean of the time for our dataset of paper based system.

$$\begin{aligned} AM_{ps} &= (12.35+6+9.56+10+20.34+9.27+9.35+6.5+11.07+8.5)/(5+2+4+4+6+4+4+2+5+3) \\ &= 102.94 / 39 \\ &= 2.6394 \end{aligned}$$

To calculate the average time of prototype digital system, we denoted the arithmetic mean as AM_{ds} . Let the time for each family be t and number of total member of all the families be n ,

$$AM_{ds} = (t_1 + t_2 + t_3 + \dots \dots \dots + t_{10}) / (n_1 + n_2 + n_3 + \dots \dots \dots + n_{10})$$

Putting the values from table one, we get the value of average mean of the time for our dataset of paper based system.

$$AM_{ds} = (7+3.55+6+7+13.20+5.49+6.4+3.5+7+6)/(5+2+4+4+6+4+4+2+5+3)$$

$$= 65.14 / 39$$

$$= 1.6702$$

$$AM_{ps} > AM_{ds}$$

Here we can see that the average value of taking single person data in regular paper based system is AM_{ps} is 2.6394 minute(s). But the average time in digital system is AM_{ds} is 1.6702 minute(s).

So, statistically it is proved that digital system takes less time for data input.

CHAPTER 4

EXPERIMENTAL RESULT AND DISCUSSION

4.1 Introduction

In this chapter we will discuss about our experiment result and will discuss about various pros and cons of our system.

4.2 Experimental Result

In the graph the X-axis represents the number of family members of the ten families. The Y-axis represents time in minutes to take the input of the data for the families.

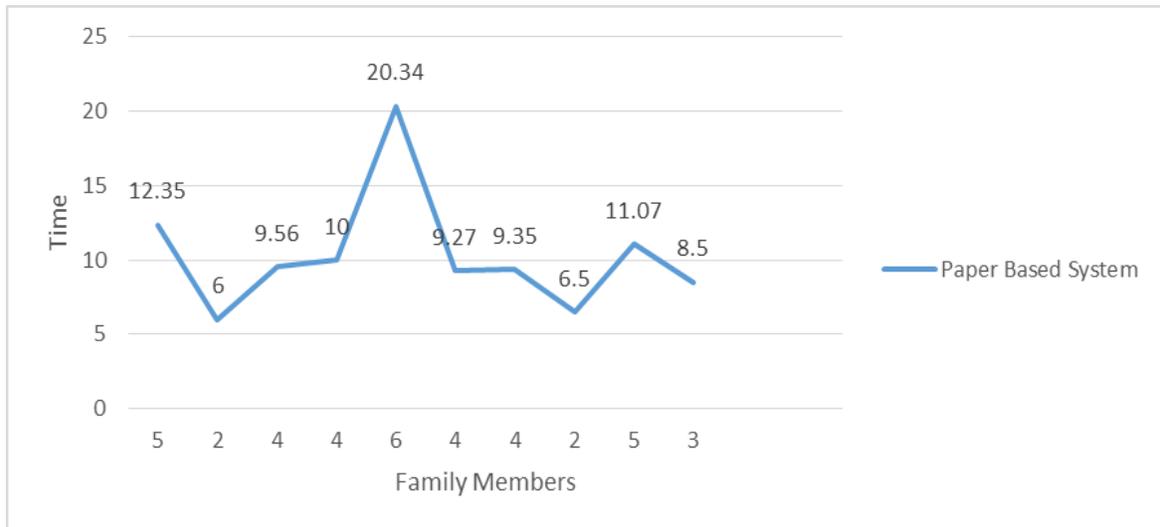


Fig 3.8: Graph of times to input data of families consisting of different members using paper based system

In the graph the X-axis represents the number of family members of the ten families. The Y-axis represents time in minutes to take the input of the data for the families.



Fig 3.9: Graph of times to input data of families consisting of different members using our designed prototype

Figure **Fig 3.8** shows that the maximum time of taking data is more than twenty minutes where from the figure **Fig 3.9** data input time is less than fourteen minutes. The two highest value of two graph has difference of almost six minutes. After merging the two graph we get a third graph **Fig 3.10** represents the time difference of two system. Here the X-axis represents the number of family members of the ten families. The Y-axis represents time in minutes to take the input of the data for the families.



Fig 3.10: Graph of differences between data input time of regular paper based system and using our prototype of digital census system

The third line graph is showing the regular system using the blue line and using our prototype is the violet line. It is clearly visible that the violet line goes under the blue one.

4.3 Descriptive Analysis of Improve Efficiency

From the two processes we gathered data and by analysis them we achieved that data gather is easier and more appropriate in the digital system. In our experiment we have added another feature of “New Information”. From the figure of the datasheet of paper based system we can visualize that if any new data comes in future, there is no space to allocate them. We faced a situation that in three families we were in in need of new information. In our digital system prototype it is a very easy process to add new information field. As the time for input data is decreasing we can use less workforce to gather more data.

From the statistical analysis we got two average time value of two different systems. The average data input value for a person is lower in the digital system. From our background study of the full process we saw that paper based process has multiple level of physical document transfer. Whereas there is no physical document transfer in digital system. So,

here time and expense will be minimized. Thus efficiency of the whole system will be increased.

4.4 Descriptive Analysis of Data Availability

From this study we observed that census digitization will be helpful if gathered data can be used for research. Customization in arranging data will add a great value to the web application. Governmental or non-governmental organization can use these data under admin authorization. Thus this system can improve data availability.

CHAPTER 5

CONCLUSION

5.1 Pros and Cons

Implementation of new system will have some pros and cons. Our system also has some pros and cons. Our system is a paperless system, no need of physical document transportation and the accuracy rate is more than paper based system.

The cons of the implementations can be described as Use of computer based mobile device may raise the cost. Device maintenance and training the users may take long time to start the system.

5.1 Future Research Plan

Digitization in census system can be a very powerful tools for storing various categorized data of the population of a country. If images, finger prints or retina scanning feature is added in this system, then the system will be helpful for security forces such as police, army and detectives. The micro data we get from this system will help us to make research more accurate ^[5].

Required instruments for these features are very expensive. We have to work on it to find some less expensive way to include these special features.

5.2 Conclusion

From this study we observed that census digitization will be helpful if gathered data can be used for research. Customization in arranging data will add a great value to the web application. This can be even more useful if we can get the finger print and the image of the people. It will be our future adding to this web application.

Reference:

- [1] *Bureau of Statistics, Bangladesh* [<http://www.bbs.gov.bd/>]
- [2] *Principles and Recommendations for Population and Housing Census.* [https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/Population-and-Housing-Censuses/Series_M67Rev2-E.pdf]
- [3] *List of national and international statistical services.* [https://en.wikipedia.org/wiki/List_of_national_and_international_statistical_services]
- [4] Swaziland goes digital for 2017 census [<http://esaro.unfpa.org/en/news/swaziland-goes-digital-2017-census>].
- [5] Ruggles, Steven. "Big microdata for population research." *Demography* 51.1 (2014): 287-297.