Advances In Multidisciplinary & Scientific Research
A Multidisciplinary & Interdisciplinary Journal

Vol. 2 No. 3, September 2016

An Enhanced Generic Data Flow and Process Control Model for Improving the Quality of Organisational Information Product

Akpon-ebiyomare D.E, Chiemeke S.C. & Egbokhare F.A
Department of Computer Science
University of Benin
Benin City, Nigeria
Email: ebiyomare77@yahoo.com;

ABSTRACT

Data and information are key assets of most organisations. Information generated by the computer system is the product of information manufacturing processes. Studies have indicated that the quality of the output from these processes depends to a large extent on the quality of the data that enters the computer system for processing by as much as 85%. Of all the data quality issues, a high percentage is attributed to human and process error. And the source of the data is a major culprit. The study empirically examined the data quality issues caused by deficient processes of process owners at the point of data creation and capture. The study identified the deficiencies in the data capture processes in Nigerian organisations and developed an enhanced generic data flow and process control model for organisations to improve the quality of their information product. The study used semi structured interviews to gather data from data producers on the data capture processes in three public sector organisations in Nigeria that have representations in all states of Nigeria. Both quantitative and qualitative methods were used to analyse the data gathered. The outcome of the analysis is a re-engineered data flow and process control model which organisations can use to ensure that all critical processes are followed for an improved quality of database product.

Keywords: Data quality, data quality improvement, process enhancement, data producers

Aims Research Journal Reference Format:
Akpon-ebiyomare D.E, Chiemeke S.C. & Egbokhare F.A (2016): An Enhanced Generic Data Flow and Process Control Model for Improving the Quality of Organisational Information Product. Advances in Multidisciplinary Research Journal. Vol. 2. No. 3, Pp197-212

1. BACKGROUND OF STUDY

High quality information is not only important to organisations but to our everyday existence. In recent times, more and more organisations are realising the importance of quality data and are therefore adopting the use of information technology to improve their data production processes. The overall objective is to improve the quality of their reports to the satisfaction of the consumers. Achieving consistent high level of quality information product entails much more than accumulation of state of the art information technology equipment (Strong et al., 2004)

They stressed that in order to assure that an organization's data is of a high standard, there is the need to focus on the quality of the data that enter their information system. Managing the quality of data would involve managing the processes that create and capture data as well as the processes that produce information. Quality data is data that is fit for purpose. Based on the study by (Chiemeke and Akpon-Ebiyomare, 2015) quality data is data that accurate, timely, and available. In addition there should be high level of responsiveness on the part of the stakeholder responsible in the organisation to make these happen.

2. STATEMENT OF PROBLEM

In order to remain competitive, organisations adopt information technology to enhance their processes with the objective of producing reports that are accurate as well as timely for the two categories of data consumers (internal and external).Quality output had always been a critical need of organisations of all kinds. This is more critical as the amount of information available for capture is rapidly on the increase. Reports indicate that data quality issues are common in many organisations and the source of the data is the culprit in in as high as 85 percent of the cases. Ballou and Pazer (2003) stated that data entry and acquisition are more error prone to both simple and complex errors. Marcus et al (2001) said that to reduce errors, a lot of effort should be put into improving this front-end process. An analysis of the requirements for a data quality improvement programme found out that the data quality researchers and practitioners, including Chiemeke and Akpon-Ebiyomare (2011); English (2009); and Loshin (2009), agree that the cause of poor quality data is often found to be human or process error. Findings from the study by Chiemeke and Akpon-Ebiyomare (2015) indicated that data consumers are not satisfied with the quality of output generated by Nigerian organisations information systems. Chiemeke and Akpon-Ebiyomare (2015) identified four critical dimensions of data quality which data consumers use to judge the quality of the information they consume. The quality of reports generated by Nigerian organisations fall short of the expectations of data consumers. Based on these, this study set out to develop a process based data quality improvement model with focus on data quality issues at the point of data creation and data capture. This is to assist organisations to monitor the activities of the front end processes with the objective to improve the quality of information system product.

3. OBJECTIVE OF STUDY

The objective of this study was investigate the data production processes of Nigerian organisations at the point of data creation and capture; develop a data flow and process control model that is focused on the front end processes and the process owners who are the data producers (those who create or collect data).

4. MATERIALS AND METHODS

This study aimed to improve information system data quality through the enhancement of the processes that code and capture the data. Process management practice is concerned with optimizing the processes with regards to enhanced quality (Zu et al., 2008). The study used face-to-face semi-structured interview method to gather data and this method gave a deep understanding of the organisations operational data environment; drew out the peculiar experiences, knowledge and perspectives of each data producer; and understanding of the activities that capture data in the organisations. The data collected from the interviews were used to develop process maps of data entry activities using flowcharting.

The activities mapping provided more insight into:

i. Employee (process owner) and data consumer involvement in each step
ii. The I.T. (information technology) system or the documents involved
iii. The individual steps involved in the data capture
iv. Data flows within the data capture, and production processes.
v. Controls in place to ensure data quality.

The fieldwork of the study was conducted both within the organisations' premises and their field locations. The population selected for this study, is made up of those that employ Information Technology (I.T.) in the capture of data in their organisations in Nigeria. They are identified by Strong et al (2004) and Wang et al (2006) as the data producers (those who create or collect data). The reason for focusing on data producers is that 85% data quality issues has been reported to come from this category of data quality stakeholders.

The data producers interviewed were selected from three organisations that collect data from Nigerians for the purpose of generating some form of identity card. The interview questions covered both close-ended and open–ended questions. The close ended questions in the interview required a 'Yes', 'No', 'Partially' or 'Dont know'.

The open ended questions provided respondents the opportunity to express their opinion on the nature of their work and the reason why certain things are done they way they are done even when it is obvious that they negatively affect quality. Based on Cronemyr (2007), the tools employed in the analysis the data and development of the model are Problem solving methodology (DRIVE), Process mapping, Process flowcharting, Data flow diagram, Checklists and Brainstorming. The respondents were interviewed individually. Table 1 represents the spread of interviewees per organisation.

Table 1: Basic Background Information about the Interview Participants.

Organisation	Gender	Total interviewed	Percentage
Organisation A	Male - 6 Female - 3	9	44.5
Organisation B	Male - 4 Female - 1	5	33.3
Organisation C	Male - 2 Female - 2	4	22.2
Total	18	18	100

5. RESULT

In order to identify the organisational data capture processes that produce data and find out where among the set of activities of the processes the data quality problems are created, the study first modelled the procedures applicants follow to obtain the information product for each of the three case organisations. These graphical representations of workflows of stepwise activities and actions are necessary for a better understanding of the processes existing in the organisations under investigation (Loshin 2009). They are shown in Figures 1, 2 and 3 for organisations A, B and C respectively.

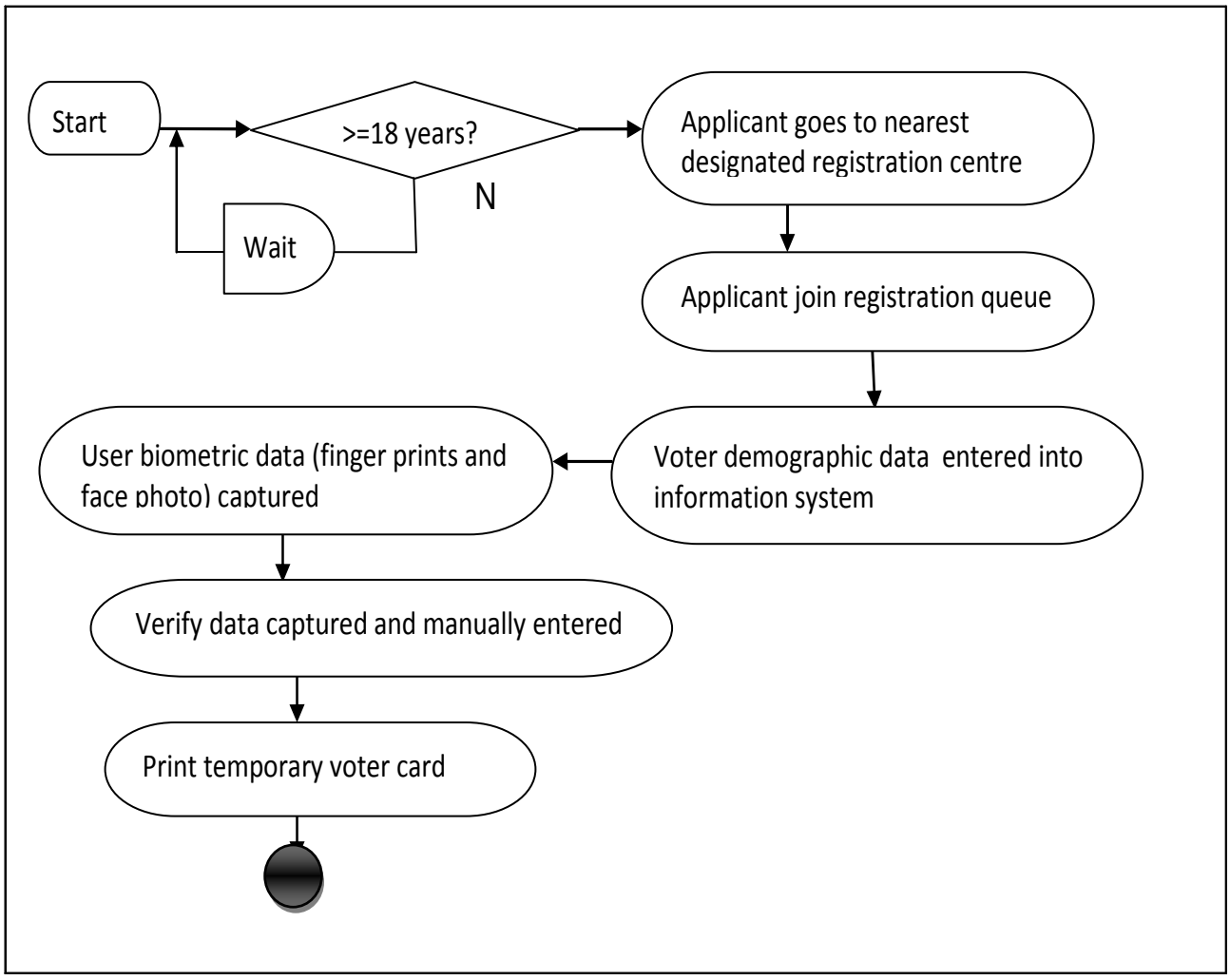

Figure 1: Data Quality Activity Map to Obtain Information System Report for Organisation A

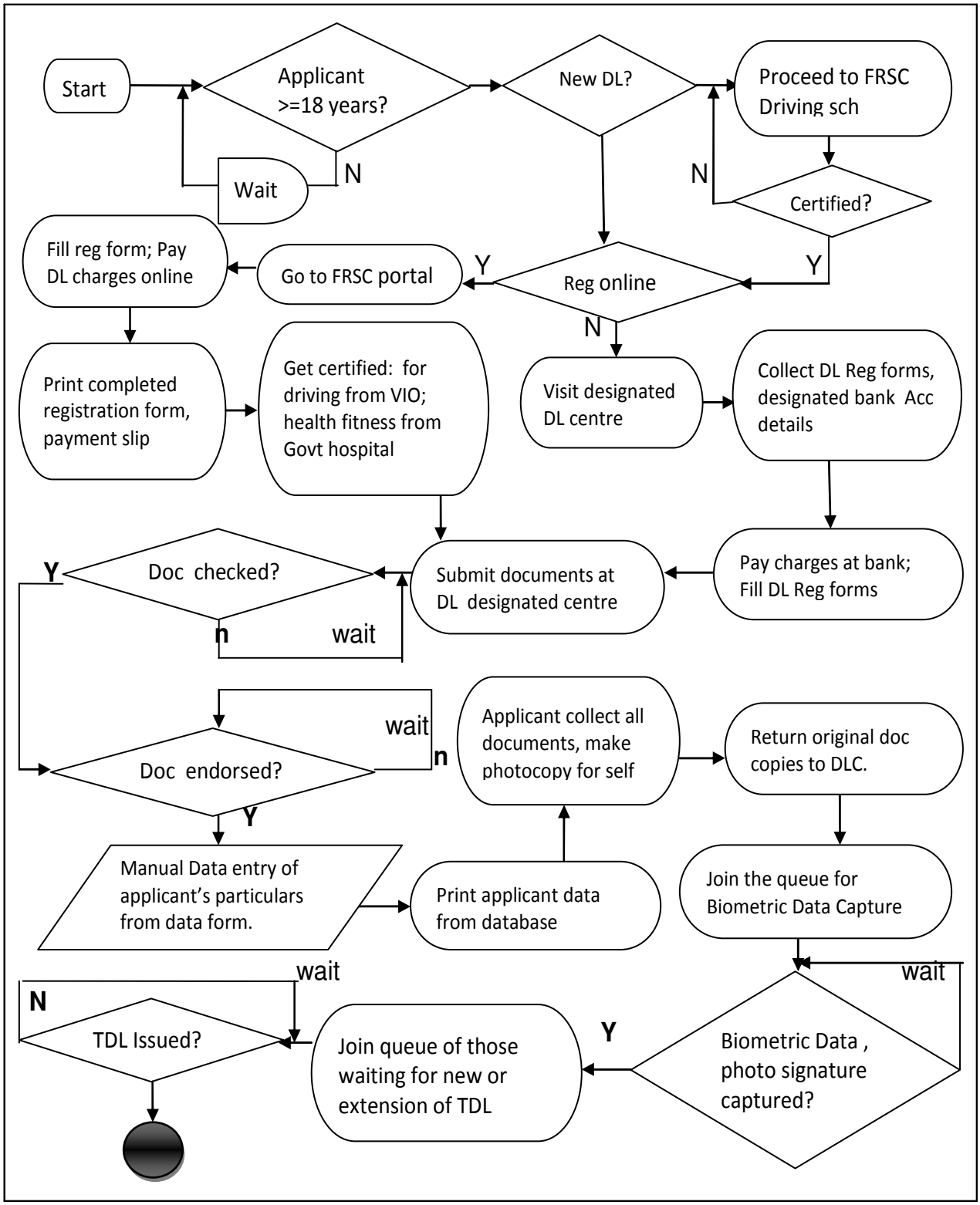

Figure 2: Data Quality Activity Map for Organisation B

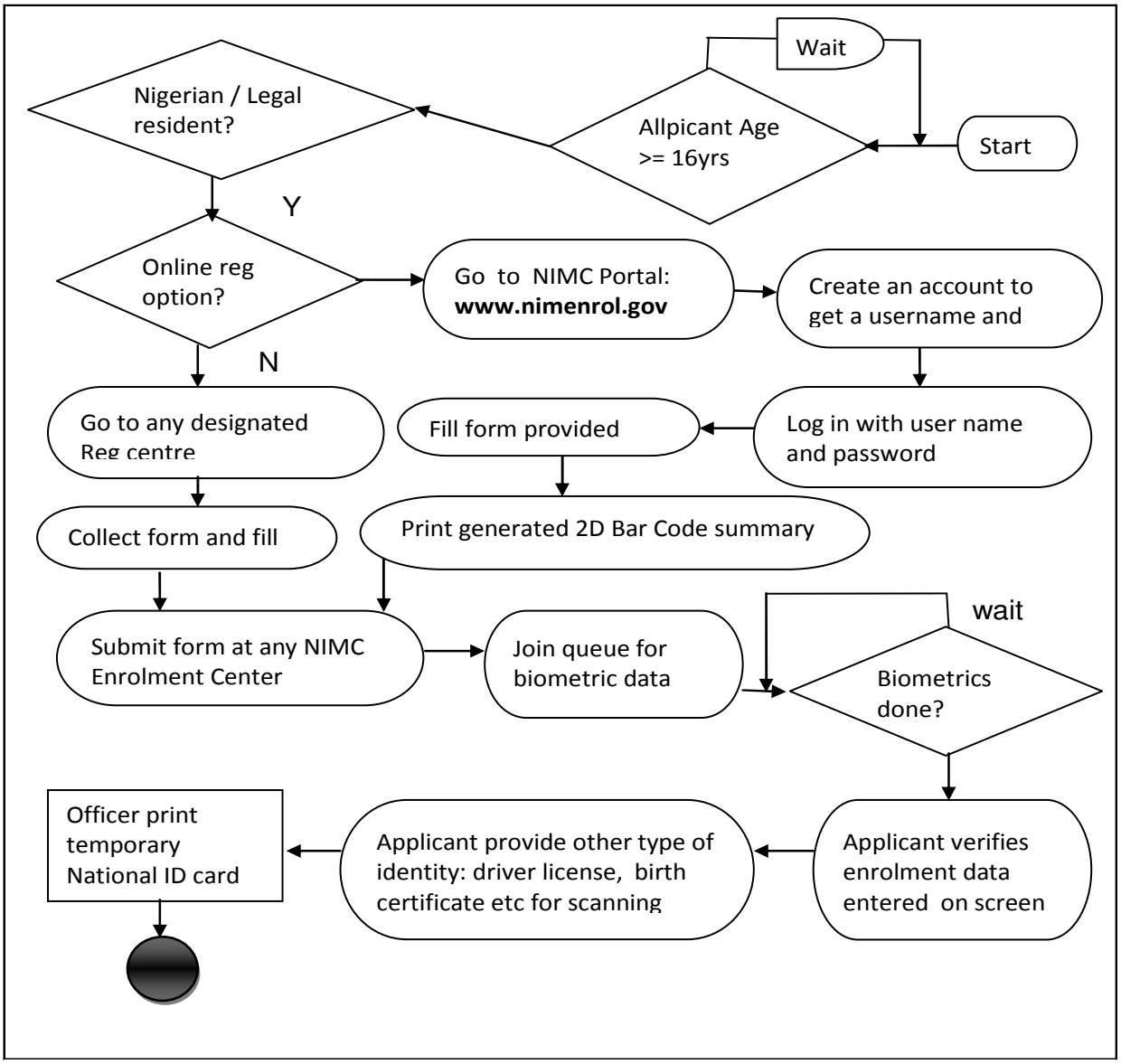

Figure 3: Data Quality Activity Map for Organisation C

5.1. Data Life Cycle

From the data quality activity map, the processes that handled data were identified and extracted in order to model the flow of data and determine the presence, deficiency or absence control. Consideration was given to those that impacted data definition, control, and data transformation. The data life cycle are modelled in Figures 4, 5 and 6 for organisations A, B and C respectively. The modelling of the controls and the processes for the three public sector organisations gave a clearer view of the activities based on the data gathered from the interviews and after interview discussions with data consumers.

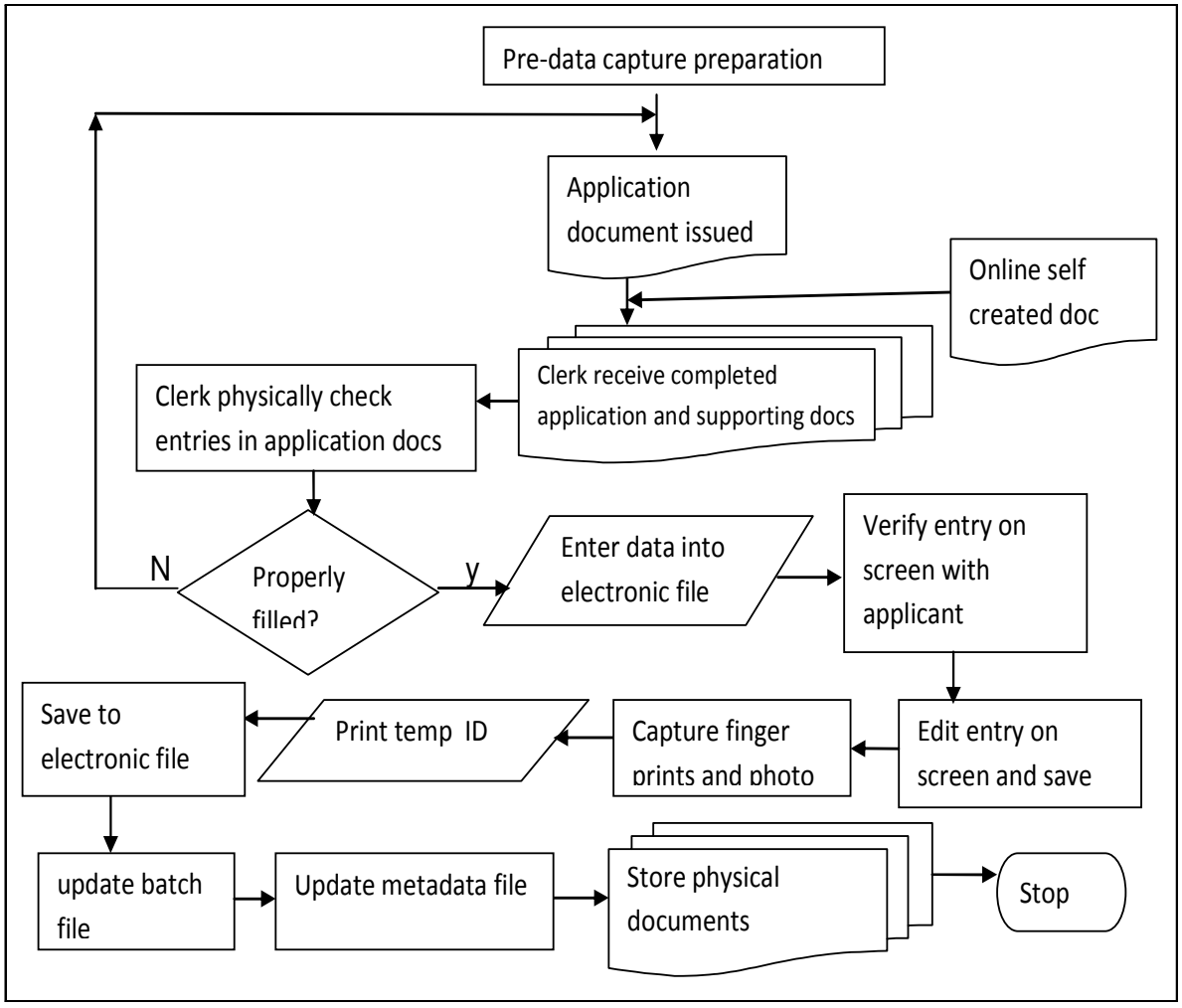

Figure 4: Data Flow Diagram for Organisation A

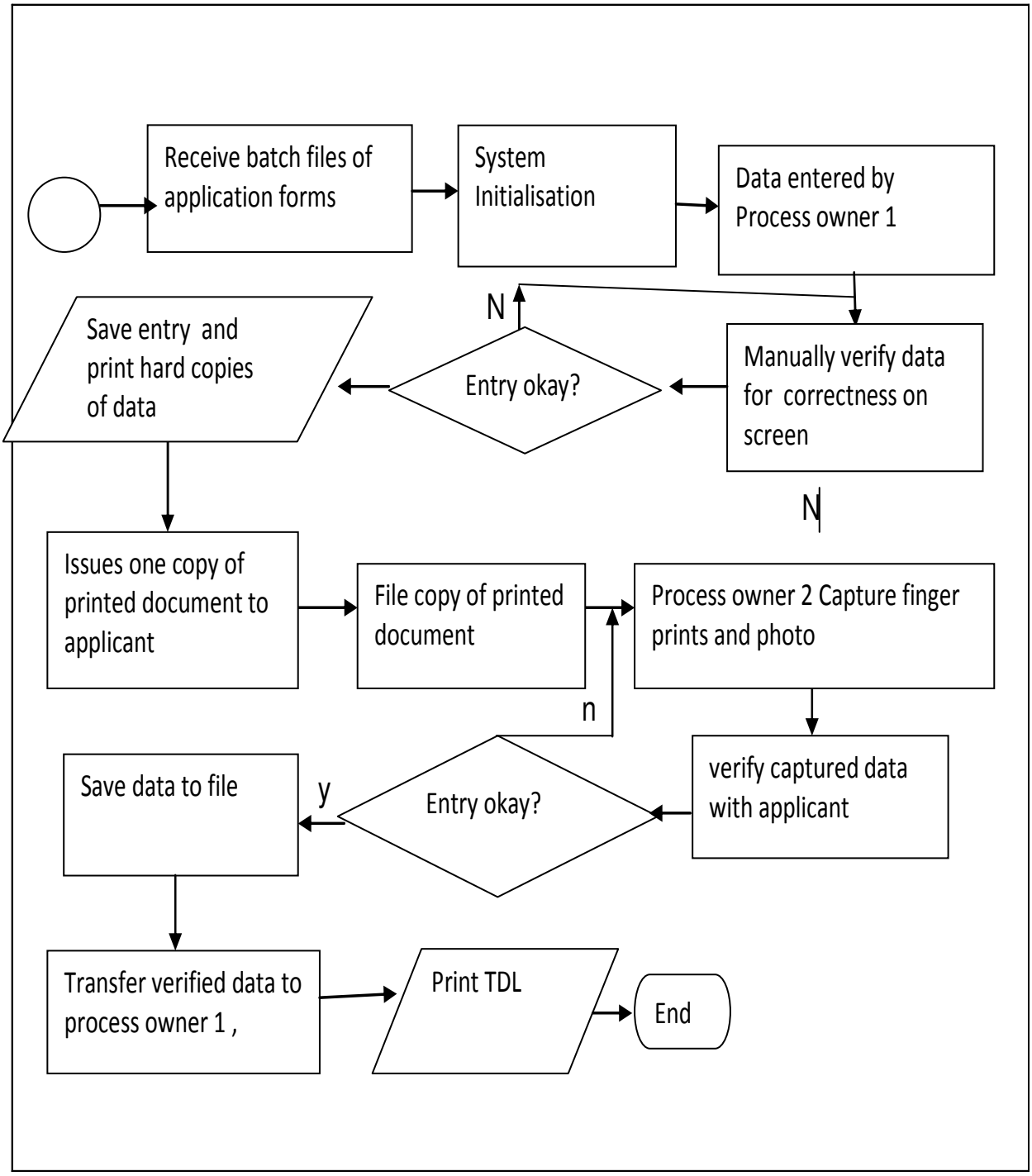

Figure 5: Data Flow Diagram (Organisation B)

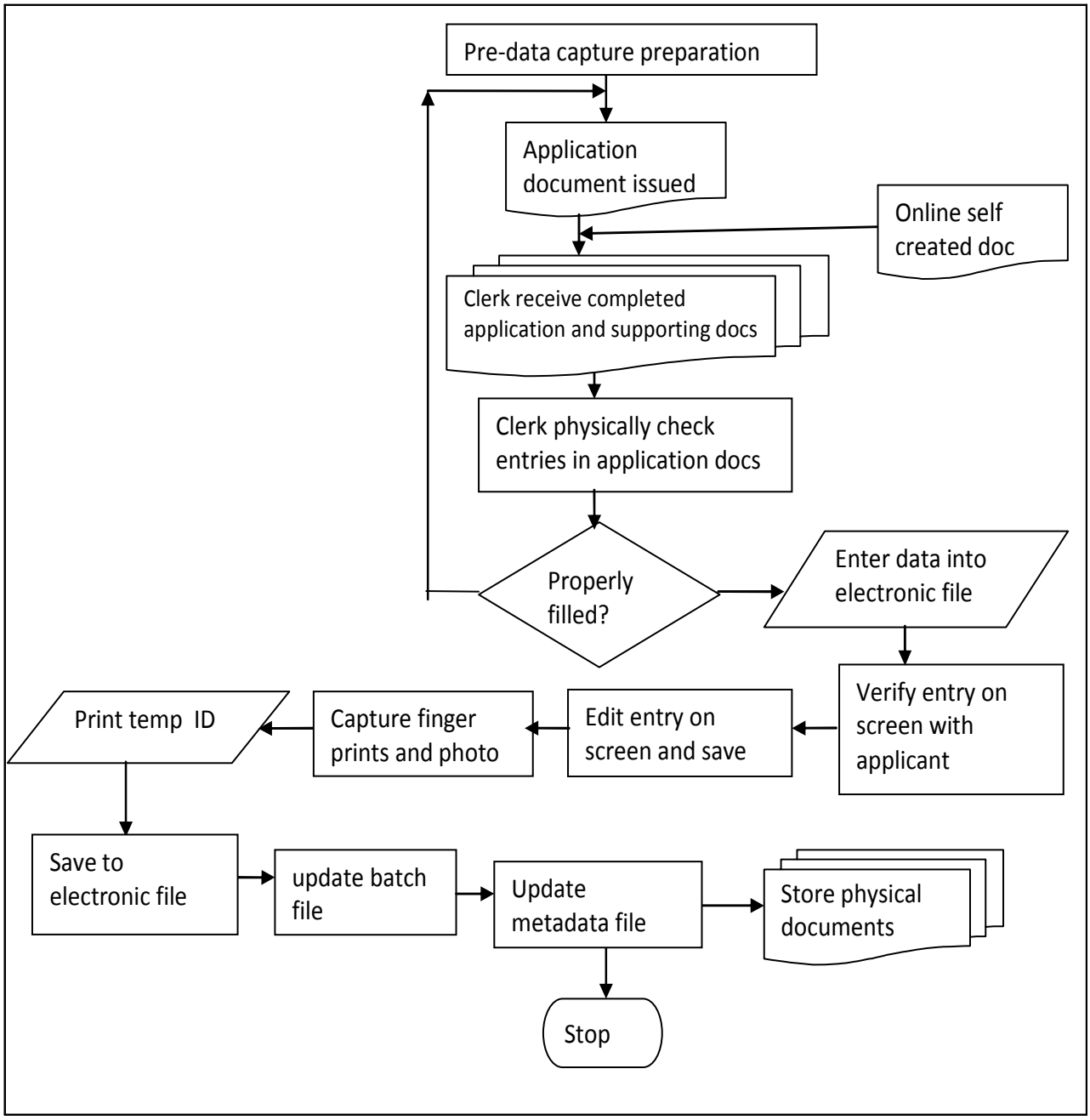

Figure 6: Data Flow Diagram (Organisation C)

5.2 Summary of Processes and Controls

The data flow diagram shows each significant step when information is created, collected, transferred, stored and deleted. The areas that cause the information quality problems are then identified in the information flow map. Figure 7 is the data entry/capture process and controls that are in place in the organisations under study. The highlighted symbol indicate that these activities are sometimes omitted by the process owners as a result of work overload, lack of awareness of the importance of the activity toward quality of output or poor supervision.

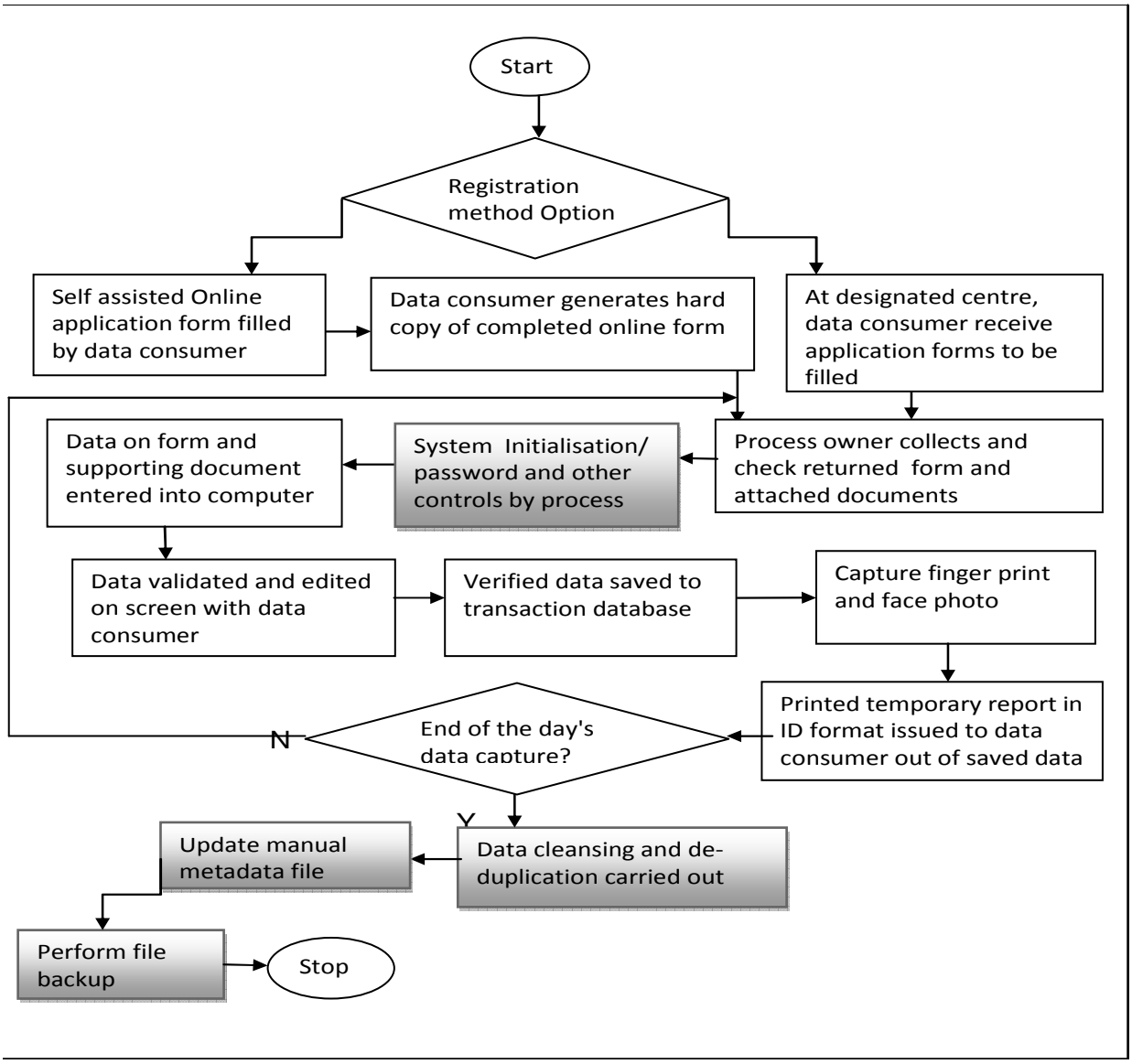

Figure 7: Integrated Data Life Cycle and Controls in Place at Organisations A, B and C

6. FINDINGS AND DISCUSSION

After modelling the controls and the processes for the three organisations to get a clearer view of the activities based on the data gathered during interviews and after interview discussions with data consumers, the problems common to the case organisations became clearer. They include:

i. There is lack of clarity around delegation of duties, authorities and data quality standards. Many times, delegation is by word of mouth.
ii. Many of the work processes are not documented
iii. Their understanding of the meaning of data quality is insufficient.
iv. There are recruitment and retention issues. Ad hoc staff are used and changed on an annual basis for data capture by one of the organisations. Many of them are not I.T. professionals. The training they receive seem insufficient for their task. Frequent change of staff affect competency.

v. Data collection processes and standards are not adequately documented
vi. Timeliness & data validation checks after data entry are limited. Sometimes this process is skipped due to the volume of data to be captured leaving workers with less time for other important processes like validation.
vii. Little attention is paid to resolving missing data or change in data consumer information in the system. There is 'change of information form' but people are either not aware or prefer re-registration to avoid the stress involved in changing information already in the database.
viii. Poor response to request for change of data consumer information already existing in the database forces data consumers to take re-registration as an option. The result of this action is duplication of records some of which go undetected.
ix. Speed of data entry is prioritized over quality of data entered. The organisations are more interested in registering as many persons as possible (and generating more money where applicable – drivers license) than in getting unique quality information of Nigerians.
x. Data is entered by personnel without sense of ownership.
xi. Absence of proper metadata (data of data) to trace owners of batches, errors or similar problems.
xii. Majority of the personnel complained of heavy workload and this is evident in the crowd of data consumers waiting to be attended to. Getting interviews from some of them meant waiting till close of work or during their short break before work resume.
xiii. Frequent equipment breakdown: Due to frequent equipment breakdown as a result of overheating, some personnel place sachets of water on their data capture equipment as a measure to tackle equipment overheating. When that fails, the system is shut down to avoid loss of registered voter data. Those waiting to register have to wait for the system to cool down. This affects timeliness.
xiv. Maintenance culture. Inadequate measures in place to prevent breakdown of equipment. Many times, equipment are not checked before being assigned for use and when they breakdown in the field, work is disrupted for the whole day as there is usually no one on ground in the field to make repairs.
xv. Absence of Nigerian name dictionary. Many Nigerian names are long and difficult to spell. The entry personnel have issues spelling them and considering the high illiteracy level of Nigerians (many cannot spell their own names) and the absence of name dictionary leaves the data entry personnel no option but to spell the names as they deem correct. Therefore the same name could be spelt differently by different entry personnel . For example an illiterate with a Nigerian name like 'Osakponmwan' or Ekhoruntonmwen' could register twice at different locations causing duplication issues that are difficult to detect as they could be spelt differently on each occassion. The issue is more complicated if the middle name of the said illiterate is equally difficult to spell (like 'Oghenerioba kpoinvwa' or 'Akporobomemerere'.).
xvi. Absence of Nigeria street name dictionary could also cause similar issues caused by name dictionary absence since most street names are Nigerian names.
xvii. Application for change of information like surname (for newly married ladies) or location address is cumbersome and in most cases not treated. Out of frustration, data consumers would rather go for a second registration because of the frustration of effecting a change in their existing record. This results in duplication of records that go undetected in most cases. The issue of de-duplication could be reduced greatly if enough attention is paid to the quality of entry and change of information.
xviii. There is yet no way of detecting that an applicant's information had done a previous registration in order to avoid double registration at point of registration/data entry. This is because there is yet no unique identification number for Nigerians. As such Nigerians are identified by their names which is in no way unique means of identification as two persons or more have been found to bear the same first name and last name.
xix. The biometric data is not yet put to full use like in detecting double application under a fake name. One person can use three different names to obtain three different ID card, driver's license or voter card without being detected despite the current registration method involving the use of biometric machines.

xx. The system is not able to detect duplicates when some information changes. Like a swap of first and middle names or using new surnames entirely and requesting for fresh license or ID or Voter card thereby creating two records of the same person. This create more work for the personnel whose duty it is to clean the database.
xxi. Scarcity of registration materials is common (eg forms for complaints).
xxii. Machine battery power used at field locations (for data capture machines) runs down resulting in disruption of work affecting timeliness dimension.
xxiii. Frequent loss of batches of data files already captured. due to poor back ups culture and poor meta data management culture.
xxiv. Corruption of data due to high temperature of machines and system breakdown
xxv. Inadequate or nonexistent provision for proper data quality assessment in-house or external.

The direct and intermediate consequences of data quality problems were identified. Each data quality issue negatively affects at least one data quality dimension. The consequences of a deficient activity in a process in which a data quality problem appears or in the absence of an important activity. A consequence can lead to several further consequences (Hill, 2009). It was found that some intermediate consequences also appeared in other activities. Each negative consequence result in a deficient data quality dimension. In line with Eppler (2006), a classification of problems based on the groupings of the dimensions they affect as well as the particular dimensions affected was developed. This method was adopted by this study as shown in Figure 8.

There after the four critical data quality dimensions identified by Chiemeke and Akpon-Ebiyomare (2015) as well as the problems that affect them are extracted shown in Figure 9 as they become the focus for data quality improvement for the case organisations. The problems that are extracted are only the ones that are related to both the data itself and the controls (Eppler, 2006). The systematic approach of mapping DQ problems to the dimensions that they affect is also in line with Lee and Haider (2011); and Strong et al, (1997).

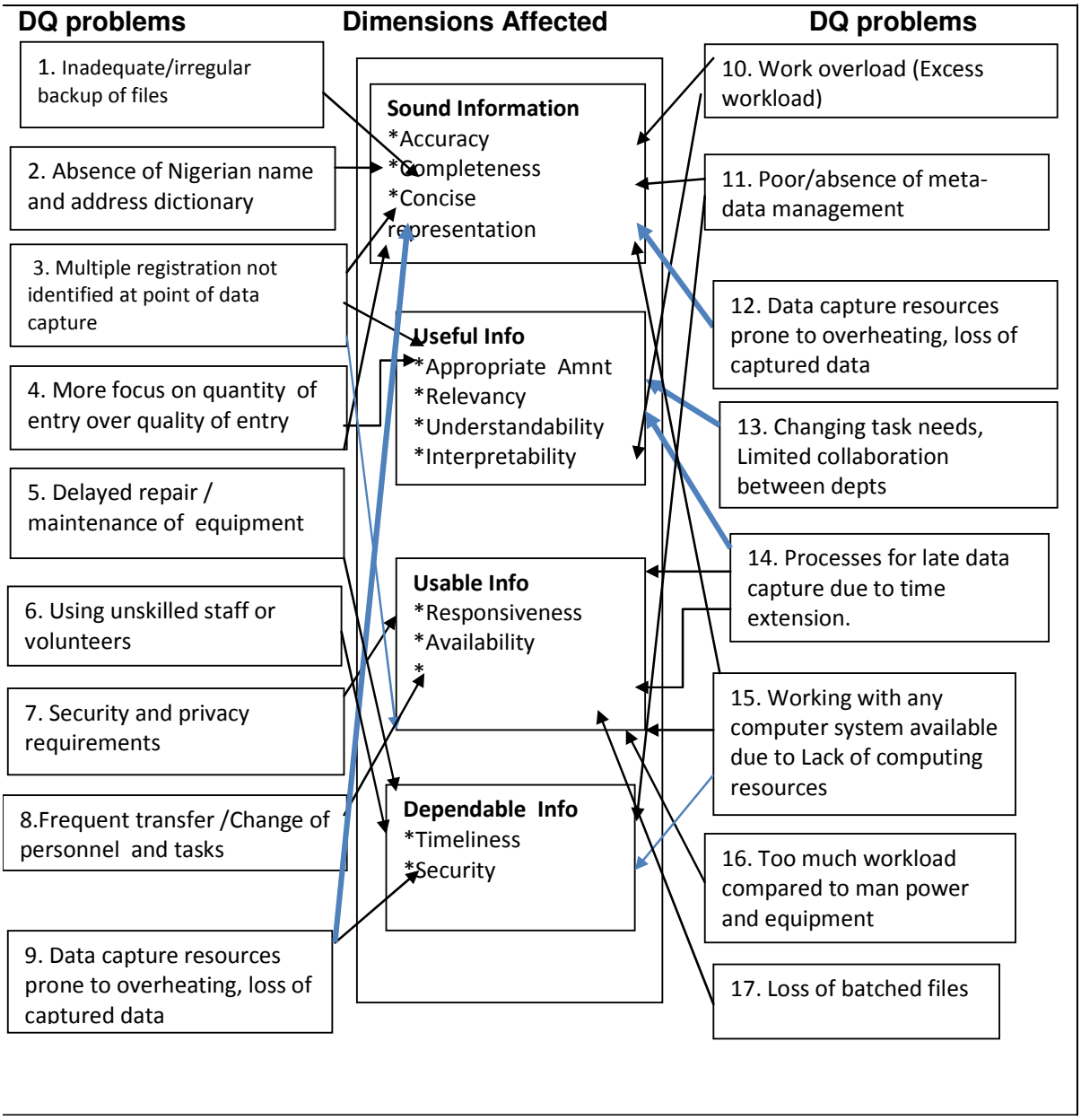

Figure 8: Mapping Data Quality Problem to Data Quality Dimensions (Based on Lee and Haider (2011).

For the purpose of clarity, the problems that directly affect the four critical data quality dimensions are the focus of this study because these are the most important dimensions to data consumers and they use them to judge the quality of data. The four dimensions were extracted from Table 8 and shown in Table 9.

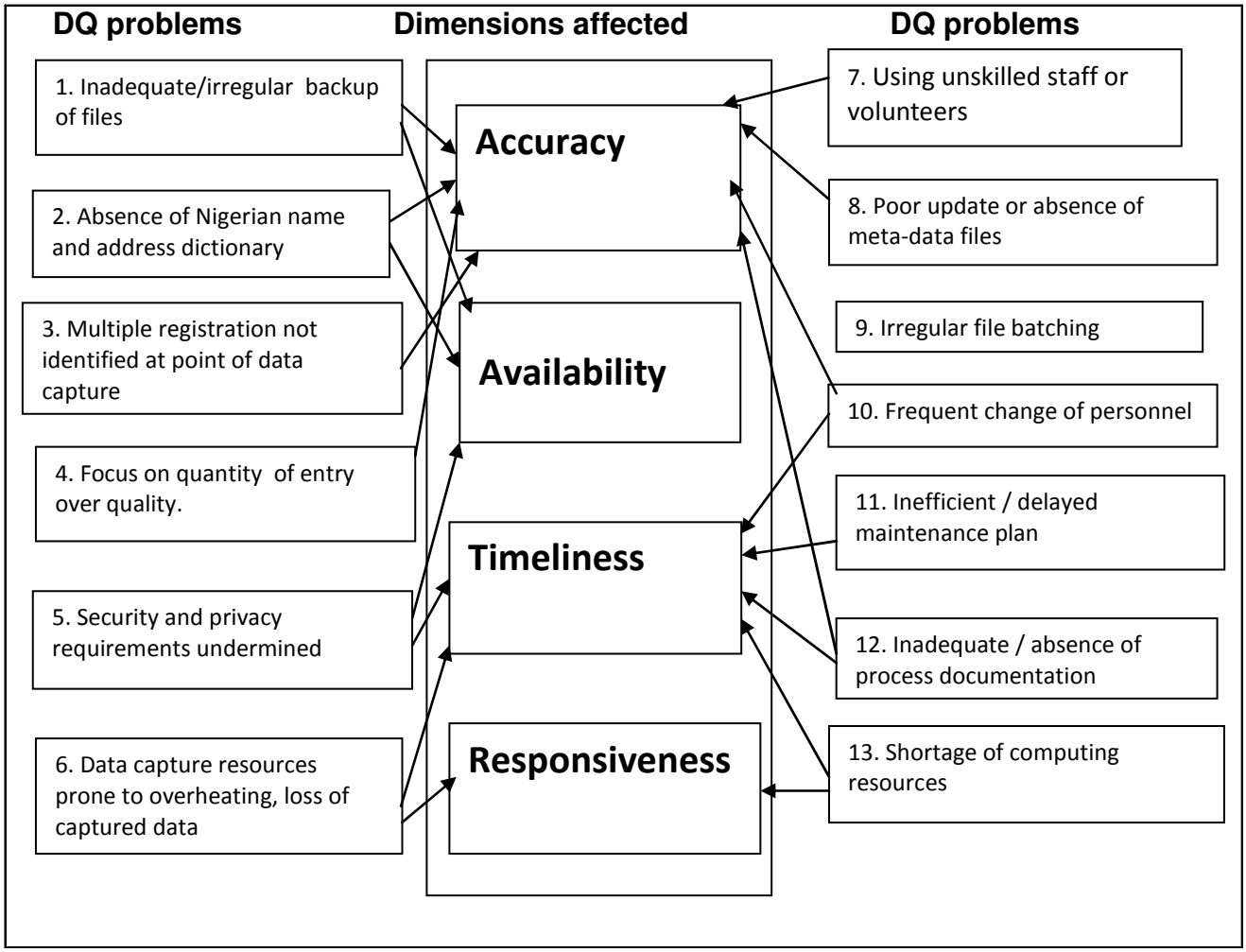

Figure 9: Mapping of The DQ Problems to The Four Critical Dimensions (Based On Lee And Haider, 2011).

Based on the findings of the analysis of the interview data, a data flow and process control model was developed (Figure 10)

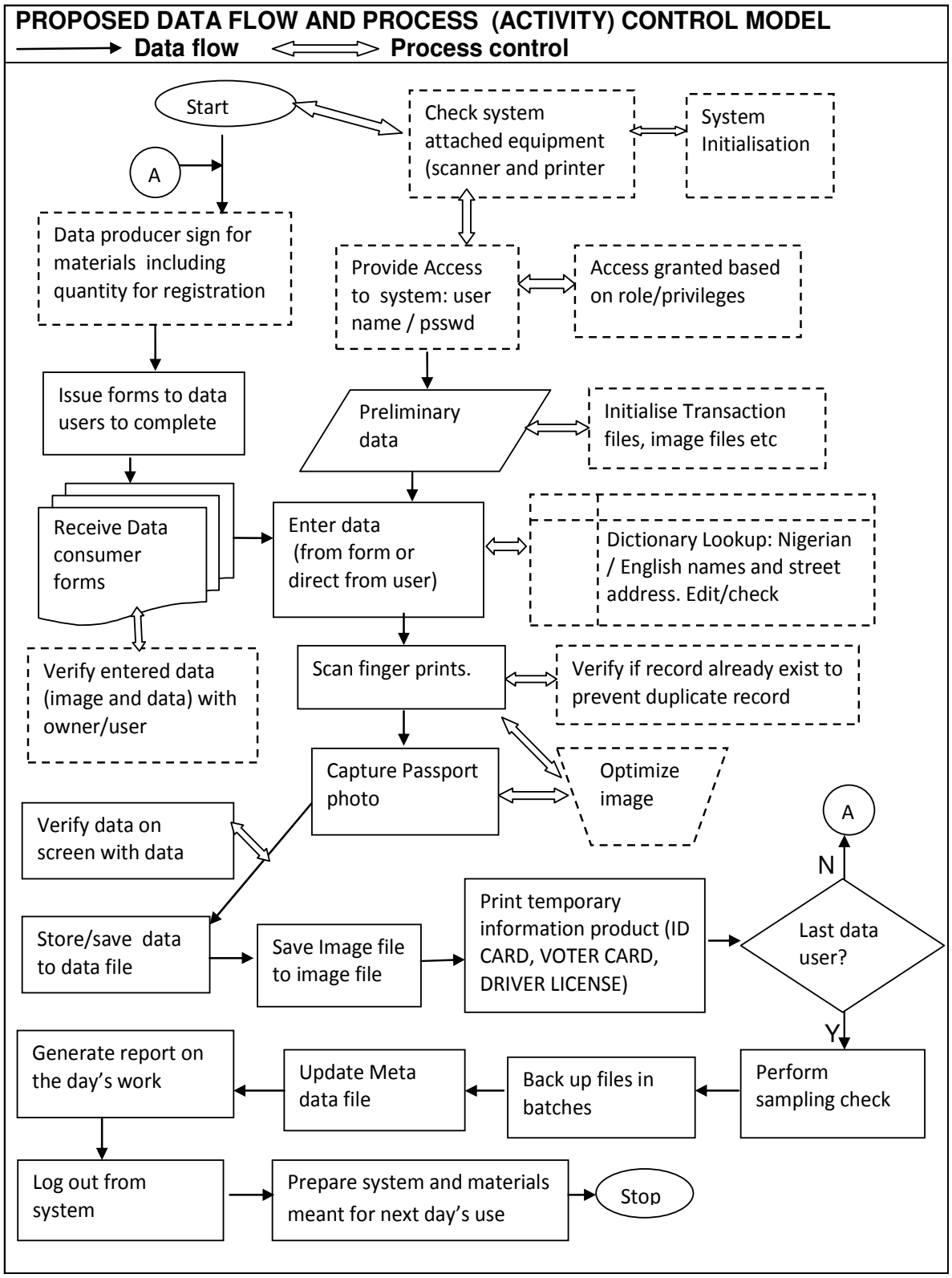

Figure 10 : Data Flow and Process Control Model (Developed for this Study)

7. CONCLUSION

The objective of this study was to develop a system for data quality improvement by improving the flow of data and incorporating required controls that are supposed to be in place to act as checks for the different critical processes.This study has shed some light on the activities of organisations and the problems faced by data producers which negatively influence the quality of data. Most of these problems handled are either process oriented or human error. The study looked at how these issues can be corrected by proposing a model of the flow of data and processes. Although results and findings from the case study are limited by the relatively small sample size and the fact that the study only considered three organisations, it does provide an initial exploratory idea about the feasibility of the proposed data quality improvement approach and this can be used to draw conclusions

6. SUGGESTIONS FOR FUTURE RESEARCH

This study developed a model that data managers can use to control the flow of data and ensure that critical processes that support high quality data are in place and in practice. The model was developed based on findings from three organisations in Nigeria. The organisations used as case study are large organisations that provide services to millions of Nigerians. For further research, a possibility would be the testing of the model using more organisations that would include small and medium scale organisations in order to provide greater understanding of the behaviour of the model under different size organisations. Another possibility for further research would be to conduct comparative studies in countries with similar characteristics as Nigeria.

REFERENCES

1. Ballou, D. P., & Pazer, H. (2005). Enhancing data quality in data warehouse environments. Communications of the ACM CACM Homepage archive. Volume 42 Issue 8. ACM New York, NY, USA
2. Cronemyr P. And Danielsson M (2007). 'Process Management 1-2-3: A Maturity Model And Diagnostics Tool. Journal Article. Total Quality Management And Business Excellence, 24(7-8) 933-944.
3. Chiemeke S.C. And Akpon-Ebiyomare D.E. (2015). 'Assessment Of The Database Product Quality Of Nigerian Organisations'. In Proceedings Of The 4th International Conference on Multidisciplinary Innovations & Technology Transfer (Mintt). Held in August, 2015 at the University Of Benin, Benin City, Nigeria.
4. Chiemeke S.C. And Akpon-Ebiyomare D.E. (2011). "High Quality Database Content: Success Factors and Improvement Strategy. In Proceedings of the 11th International Conference of the Nigerian Computer Society (NCS). Held in Abuja, Nigeria.
5. English, L (2009). Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits. New York, NY: Wiley Computer Publishing.
6. Eppler, M. J. (2006). "Managing information quality: increasing the value of information in knowledge-intensive products and processes (2nd Edition ed.). "Berlin: Springer.
7. Lee S.H. and Haider A. (2011). "A Framework for Information Quality Assessment Using Six Sigma Approach". Vol. 2011, Article ID 927907 IBIMA Publishing Communications of the IBIMA
8. Loshin D. (2009) "Master Data Management" Oxford Elsevier Inc.
9. Olson, J. E. (2003). Data Quality. The Accuracy Dimension. San Francisco: Morgan Kaufmann Publishers.
10. Strong, D., Lee Y., and Wang. R. (2004), the life of data quality projects. Working paper, MIT TDQM Research Program, E53-320, 50 Memorial Drive, Cambridge, 02139.
11. Wang R.Y., Lee Y.W., Pipino L.L., Funk J.D.(2006). Journey to Data Quality. The MIT Press .
12. Zu, X., Fredendall, L., & Douglas, T. (2008). The Evolving Theory of Quality Management: The Role of Six Sigma. Journal of Operations Management, 26, 630 – 650.