

**A TAXONOMY OF DATA BREACHES: DEFINING THE IMPACT OF DATA
BREACH AND UNDERSTANDING ITS THREAT TO THE UNITED STATES
GOVERNMENT FACILITIES**

EMEJOR, ONORIODE BRYAN

(19PBE01927)

B.sc Mass Communication, Bowen University, Iwo.

SEPTEMBER, 2021

**A TAXONOMY OF DATA BREACHES: DEFINING THE IMPACT OF DATA
BREACH AND UNDERSTANDING ITS THREAT TO UNITED STATES
GOVERNMENT FACILITIES**

BY

**EMEJOR, ONORIODE BRYAN
(19PBE01927)**

B.sc Mass Communication, Bowen University, Iwo.

**A DISSERTATION SUBMITTED TO THE SCHOOL OF
POSTGRADUATE STUDIES IN PARTIAL FULFILMENT OF THE
REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE
(M.Sc.) DEGREE IN MASS COMMUNICATION IN THE
DEPARTMENT OF MASS COMMUNICATION, COLLEGE OF
MANAGEMENT AND SOCIAL SCIENCES, COVENANT
UNIVERSITY, OTA.**

SEPTEMBER, 2021

ACCEPTANCE

This is to attest that this dissertation is accepted in partial fulfilment of the requirements for the award of the degree of Masters of Sciences in Mass Communication in the Department of Mass Communication, College of Management and Social Sciences, Covenant University, Ota, Nigeria.

Mr. John A. Philip

(Secretary, School of Post-Graduate Studies)

.....

Signature and Date

Prof. Akan B. Williams

(Dean, School of Postgraduate Studies)

.....

Signature and Date

DECLARATION

I, **EMEJOR, ONORIODE BRYAN (19PBE01927)** declare that this research was carried out by me under the supervision of Dr Ada Sonia Peter, Department of Mass Communication, College of Management and Social Sciences, Covenant University, Ota, Nigeria. I attest that the dissertation has not being presented either wholly or partially for the award of any degree elsewhere. All sources of data and scholarly information used in this dissertation are duly acknowledged.

EMEJOR, ONORIODE BRYAN

Signature and Date

CERTIFICATION

We certify that this dissertation titled **“A TAXONOMY OF DATA BREACHES: DEFINING THE IMPACTS OF DATA BREACH AND UNDERSTANDING ITS THREAT TO THE UNITED STATES GOVERNMENT FACILITIES”** is an original research work carried out by **EMEJOR, ONORIODE BRYAN (19PBE01927)** in the Department of Mass Communication, College of Management and Social Sciences, Covenant University, Ota, Ogun State, Nigeria under the supervision of Dr. Ada Sonia Peter. We have found this work acceptable as part of the requirements of the award for Master of Science (M.Sc.) in Mass Communication.

Dr. Ada Sonia Peter
(Supervisor)

.....
Signature and Date

Dr. Kehinde Oyesomi
(Head of Department)

.....
Signature and Date

Prof. Adepoju Tejumaiye
(External Examiner)

.....
Signature and Date

Prof. Akan B. Williams
(Dean, School of Postgraduate Studies)

.....
Signature and Date

DEDICATION

This dissertation is dedicated to God Almighty for His goodness and mercies. Also, this work is dedicated to my parents and to all publishers, researchers, students, editors, authors and journalists.

ACKNOWLEDGEMENTS

My appreciation goes to God Almighty, the author and finisher of my faith. For His grace in sustaining me throughout my programme, for his unending faithfulness, for bestowing understanding unto me, I am grateful. I am also grateful to the founder of Covenant University, Dr David Olaniyi Oyedepo for his Godly leadership.

My thanks go to the management under the able leadership of Professor Abiodun H. Adebayo (Vice-Chancellor), Pastor Oluwasegun Omidiora (The Registrar) and Professor Uwalomwa Uwuigbe (Dean, College of Management, and Social Sciences). Also, I thank the Dean, School of Postgraduate Studies, Professor Akan B. Williams, and the Sub-Dean, Dr. Emmanuel Amoo, for their leadership. The selfless service of the Head, Mass Communication Department, Dr Kehinde Oyesomi is deeply appreciated. May the good Lord bless you and reward you for the knowledge you imparted in me.

My heartfelt appreciation also goes to Dr. Ada Sonia Peter, who oversaw this work to completion. I really want to say thank you for believing in me, for your sacrifice and humility, for showing me by example that nothing is impossible to achieve. You're a rare gem. Thank you very much Ma. I also want to acknowledge my lecturer, Dr Oscar Odiboh, for always encouraging us all and for your interest in our progress. To my course mates- Kourtney Sunday, Omokiti Fegor, Nejo Elizabeth, Mercy Banda, Aiyende Faith, Salau Adejonwo and Levi Freeman. This journey was beautiful with you guys in it. To my friends, Paul David, Hannah Ohore- (thank you so much for being a wonderful friend indeed and for all the sleepless nights you had to do). Most importantly, I am grateful for my support system, my parents and siblings. I can't thank you enough for your encouragement, spiritual and financial support. For those I could not mention, you are acknowledged and I am thankful for the part you all have played in this academic journey.

I acknowledge you all. Thank you so much for your care and love. God bless you dearly.

TABLE OF CONTENTS

COVER PAGE	PAGE
TITLE PAGE	ii
ACCEPTANCE	iii
DECLARATION	iv
CERTIFICATION	v
DEDICATION	vi
ACKNOWLEDGEMENTS	vii
TABLE OF CONTENTS	viii
LIST OF FIGURES	xii
ABSTRACT	xiii
CHAPTER ONE	
INTRODUCTION	1
1.1 Background to the Study	1
1.2 Statement of Research Problem	3
1.3 Objectives of Study	4
1.4 Research Questions	5
1.5 Significance of study	5
1.6 Scope of study	6
1.7 Limitations to the study	6
1.8 Operational definition of terms	6
CHAPTER TWO	
LITERATURE REVIEW	8
2.0 Preamble	8
2.1 The Conceptual review	8
2.2 Data breach in the United States Government Facilities	11

2.3 Empirical Reviews	12
2.4 Implication of Data Breach	19
2.5 Cyber and National Security	20
2.6 Cybercrime/Data Breaches	21
2.7 Need for Taxonomy of Breaches	22
2.8 National Security	23
2.9 Forms of Security	23
2.10 Classification of Digital Theft and Breaches	24
2.10.1 Data Theft	24
2.10.2 Anatomy of a Data Breach	24
2.10.3 Privacy/Password Violation	24
2.10.4 Phishing	25
2.11 Developing a taxonomy of data breaches of US Government facilities	25
2.11.1 Taxonomy of Data Breaches	26
2.12 Theoretical Framework	33
2.12.1 The Swiss Model of Human Error	33
2.12.2 Relevance of the Swiss cheese model of Human Error	34
2.13 Gaps to fill	35
CHAPTER THREE	
METHODOLOGY	36
3.1 Introduction	36
3.2 Research Design	36
3.3 Data Collection Method	36
3.4 Data Extraction Tool	37
3.5 Population and Sample Size	38
3.6 Data Cleaning	38
3.7 Search Strategy	39
3.8 Sampling Technique	39
3.9 Labelling Data for Testing and Training the algorithms	40

CHAPTER FOUR	
DATA ANALYSIS AND PRESENTATION	41
4.1 Preamble	41
4.2 Justification of Research Questions	41
4.2.1 Research Question One	41
4.2.2 Research Question Two	60
4.2.3 Research Question Three	63
4.2.4 Research Question Four	65
CHAPTER FIVE	
DISCUSSION OF RESULTS AND FINDINGS	70
5.1 Preamble	70
5.2 Discussion of Objectives	70
5.2.1 Objective One	70
5.2.2 Objective Two	71
5.2.3 Objective Three	71
5.2.4 Objective Four	72
CHAPTER SIX	
CONCLUSION AND RECOMMENDATION	73
6.1 Preamble	73
6.2 Conclusion	73
6.3 Recommendations	74
6.4 Contribution to knowledge	75
6.5 Areas for Further Research	75
REFERENCES	77

LIST OF FIGURES

Figure 1: Annual number of data breaches and exposed records in the United States from 2005 to 2020 (in millions)

Figure 2.1: Diagram showing types of data breaches

Figure 2.2: Diagram showing categories of data breaches according to Verizon

Figure 2.3: Diagrammatic representation of taxonomy of data breaches

Figure 2.4: Risk management approach diagram

Figure 4.1.1: Chart of data breach by hacking

Figure 4.1.2: Chart of data breach by phishing

Figure 4.1.3: Chart of data breach by malware

Figure 4.1.4: Chart of data breach by unauthorized access

Figure 4.1.5: Chart of data breach by paper data

Figure 4.1.6: Chart of data breach by physical

Figure 4.1.7: Chart of data breach by ransomware

Figure 4.1.8: Chart of data breach by SQL injection

Figure 4.1.9: Chart of data breach by unknown

Figure 4.1.10: Chart of data breach in Pennsylvania

Figure 4.1.11: Chart of data breach in New York

Figure 4.1.12: Chart of data breach in Georgia

Figure 4.1.13: Chart of data breach in Arkansas

Figure 4.1.14: Chart of data breach in Virginia

Figure 4.1.15: Chart of data breach in California

Figure 4.1.16: Chart of data breach in Indiana

Figure 4.1.17: Chart of data breach in Missouri

Figure 4.1.18: Chart of data breach in Texas

Figure 4.1.19: Chart of data breach in Michigan

Figure 4.2: Heat map of the spread of data breach across The United States of America

Figure 4.4.1: Screenshot of machine learning algorithm

Figure 4.4.2: Screenshot of artificial intelligence learning

Figure 4.4.3: Screenshot of predictive algorithm

Figure 4.4.4: Screenshot of an algorithm predicting the impact

ABSTRACT

In 2020, four United States key federal agencies, from the Department of Homeland Security to the agency that oversees America's nuclear weapons arsenal to tech and security companies, including Microsoft, were breached. Weeks after the United States government announced that multiple federal agencies had been targeted, the full scope and consequences of the suspected Russian hack remained unknown. Investigators struggled to determine what information the hackers may have stolen and what they could do with it. The struggle implied a lack of the scientific framework upon which governments can swiftly identify the possible scope and consequences of the data breaches in the government facilities. Hence, while previous studies may have developed some form of a data breach or cyber harm taxonomies, this study seeks to train a machine learning algorithm that will use existing taxonomy of the prevalence, incidence, and consequences of data breaches on the United States government facilities sector to predict future consequences of similar attacks. The study used available data to capture the prevalence, incidence, and implications of the data breaches on the government facilities sector then used the same to train an algorithm (LSVM) that can provide insight to possible consequences, response, and spread of new attacks. The scope and data used for the study are limited to data breaches that occurred in the United States government facilities between the years 2000 and 2021. The outcome of this is a machine learning tool that suggests and detects probable consequences of each type of data breach. The tool will be useful for researchers and practitioners alike to consider the full range of consequences that might result from different kinds of data breaches when developing response tactics. The tool is available on Streamlit:

https://share.streamlit.io/bryanemejor/data_breach_thesis/main/Stream_Bryan.py

Keywords: *data, data breach, government-industry, hacking, phishing, Ransomware*