

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.



Archiving Social Survey Data in Africa: an Overview of African Microdata Curation and the Role  
of Survey Data Archives in Data Management in Africa

By

**Lynn Woolfrey**  
**ERSHEI002**

A Thesis Submitted in Fulfilment of the Requirements for the Degree of  
Master of Philosophy in Library and Information Science

UNIVERSITY OF CAPE TOWN

2009

SUPERVISORS

Professor P. Underwood  
Associate Professor K. de Jager

DEPARTMENT OF INFORMATION AND LIBRARY STUDIES

## PLAGIARISM DECLARATION

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.
2. I have used the **Chicago Convention** for citation and referencing. Each contribution to, and quotation in, this thesis from the work(s) of other people has been attributed, and has been cited and referenced.
3. This thesis is my own work.
4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

**Signature**

University of Cape Town

## **ACKNOWLEDGMENTS**

I would like to acknowledge those who have given encouragement during the creation of this work: David, Sean and Beata, Omar, friends and members of the Mermaid Club, with gratitude for their wisdom and tolerance.

I also wish to thank my supervisors for their thoroughness and concern, and my colleagues at DataFirst, Matthew Welch and Alison Siljeur, for their unfailing support.

University of Cape Town

## ABSTRACT

This study examines current practice in the curation of social survey data in African countries and makes suggestions for future improvements in this regard. Curation of data refers to its preservation and management for reuse.

Utilising survey data for the study of social phenomena other than those for which the original survey was initiated is a relatively new research approach in Africa. Thus best practice for this type of research is still being put in place by African organisations. This involves the development of optimal means of processing and storing the data for re-use. Of concern to this study is what constitutes the most effective way of managing and sharing the information garnered from these surveys as a resource for economic and social development in Africa.

Social survey data refers to both the statistical information which is the final product of censuses or sample surveys, and the documentation provided with the data to facilitate its reuse. Documentation includes technical notes and questionnaires used in the survey process, as well as metadata (detailed information about the data) and reports produced concerning the final survey findings.

The research looks at the history of the management of social survey data worldwide and in African countries, and the policies and processes involved in curating survey information in these countries. The comparative component of the study examines developments in this field internationally and compares these to practices on the African continent. International best practice in the field has been used to evaluate current methods of survey data archiving in African countries. The study presents strategies to ensure the optimal preservation and effective sharing of survey data among countries of the region. Strategies for the establishment of a Pan African network of data sharing organisations are suggested to support future repurposing of African census and survey data.

## LIST OF ACRONYMS

<b>AAPA</b>	Addis Ababa Plan of Action for Statistical Development in Africa in the 1990s
<b>AASDA</b>	African Association of Statistical Data Archivists
<b>ACAP</b>	African Census Analysis Project
<b>ACBF</b>	African Capacity-Building Foundation
<b>ACM</b>	Association for Computing Machinery
<b>ACS</b>	African Centre for Statistics
<b>ADB</b>	African Development Bank
<b>ADP</b>	Accelerated Data Program
<b>AERC</b>	African Economic Research Consortium
<b>AFRISTAT</b>	Economic and Statistical Observatory for Sub-Saharan Africa
<b>AFRITAC</b>	African Technical Assistance Centre
<b>AHSCP</b>	African Household Survey Capability Programme
<b>AISI</b>	African Information Society Initiative
<b>APDU</b>	Association of Public Data Users
<b>ASPA</b>	AFRISTAT Strategic Plan of Activities
<b>ASSD</b>	African Symposium on Statistical Development
<b>ASTP</b>	African Statistical Training Programme
<b>AU</b>	African Union
<b>CASD</b>	Coordinating Committee on African Statistical Development
<b>CESSDA</b>	Committee for European Social Science Data Archives
<b>CODATA</b>	Committee on Data for Science and Technology
<b>CODI</b>	Committee on Development Information
<b>CPA</b>	Comprehensive Peace Agreement
<b>CSA</b>	Central Statistical Authority of Ethiopia
<b>CSSDA</b>	Committee on Social Science Data Archives
<b>CSSR</b>	Country Statistical Situation Report
<b>CST</b>	Country Technical Services Team
<b>DA</b>	District Assembly
<b>DDI</b>	Data Documentation Initiative
<b>DMID</b>	Data Management and Information Delivery
<b>DQAF</b>	Data Quality Assessment Framework

<b>DST</b>	Department of Science and Technology
<b>EASTC</b>	East African Statistical Training Centre
<b>EC</b>	European Commission
<b>ECASS</b>	European Centre for Analysis in the Social Sciences
<b>ECOSOC</b>	United Nations Economic and Social Council
<b>ECSSID</b>	European Cooperation in Social Science Information and Documentation
<b>EDAN</b>	East European Data Archive Network
<b>ENEA</b>	National Institute of Applied Economics
<b>ENSEA</b>	National Higher Institute of Applied Statistics and Economics
<b>ESFRI</b>	European Strategy Forum on Research Infrastructures
<b>ESDS</b>	Economic and Social Data Service
<b>ESRC</b>	UK Economic and Social Research Council
<b>EU</b>	European Union
<b>EUROSTAT</b>	Statistical Office of the European Commission
<b>FASDEV</b>	Forum for African Statistical Development
<b>GDDS</b>	General Data Dissemination Standard
<b>GERD</b>	Gross Expenditure on Research and Development
<b>GSS</b>	Ghana Statistical Service
<b>HSRC</b>	Human Sciences Research Council
<b>IASSIST</b>	International Association for Social Science Information Services and Technology
<b>ICP-Africa</b>	International Comparison Program for Africa
<b>ICPSR</b>	Inter-University Consortium for Political and Social Research
<b>ICSSD</b>	International Committee for Social Sciences Documentation
<b>ICT</b>	Information and Communication Technologies
<b>ICT4D</b>	Information and Communication Technology for Development
<b>IDF</b>	International Data Forum
<b>IFA</b>	Information for All Programme
<b>IFDO</b>	International Federation of Survey Data Organisations
<b>IFORD</b>	Institute of Demographic Training and Research
<b>IFPRI</b>	International Food Policy Research Institute

<b>IHSN</b>	International Household Survey Network
<b>IMF</b>	International Monetary Fund
<b>INE</b>	Instituto Nacional de Estadística
<b>INSEED</b>	Institute of Statistics, Economics and Demographic Studies
<b>ISAE</b>	Institute of Statistics and Applied Economics
<b>ISER</b>	Institute for Social and Economic Research
<b>ISSC</b>	International Social Science Council
<b>ISSEA</b>	Sub-Regional Institute of Applied Statistics and Economics
<b>KIDS</b>	KwaZulu-Natal Income Dynamics Study
<b>LSMS</b>	Living Standards Measurement Study
<b>MADIERA</b>	Multilingual Access to Data Infrastructures of the European Research Area
<b>MAPS</b>	Marrakesh Action Plan for Statistics
<b>MDG</b>	Millennium Development Goal
<b>MEDSTAT</b>	Mediterranean Statistical Co-operation Programme
<b>NADA</b>	National Data Archive
<b>NATIS</b>	National Information Systems
<b>NBS</b>	Nigerian Bureau of Statistics
<b>NeDICC</b>	Network of African Data and Information Curation Centres
<b>NEPAD</b>	New Partnership for Africa's Development
<b>NESSTAR</b>	Networked Social Science Tools and Resources
<b>NICI</b>	National Information and Communication Infrastructure
<b>NIP</b>	National Information Policy
<b>NRI</b>	Networked Readiness Index
<b>NSDS</b>	National Strategies for the Development of Statistics
<b>NSF</b>	National Science Foundation (US)
<b>NSO</b>	National Statistics Office
<b>NSS</b>	National Statistical System
<b>OAU</b>	Organisation of African Unity
<b>OECD</b>	Organisation for Economic Cooperation and Development
<b>OHS</b>	October Household Survey

<b>ONS</b>	Office of National Statistics (UK)
<b>ORSTROM</b>	Office de la Recherche Scientifique et Technique d'Outre Mer
<b>OSS</b>	Oslo Summer School
<b>PADIS</b>	Pan-African Development Information System
<b>PARIS21</b>	Partnership in Statistics for Development in the 21 <sup>st</sup> Century
<b>PGI</b>	General Information Programme
<b>PPDS</b>	Programme Population et Developpement au Sahel
<b>PRSP</b>	Poverty Reduction Strategy Paper
<b>PSLSD</b>	Project for Statistics on Living Standards and Development
<b>RICI</b>	Regional Information and Communication Infrastructure
<b>RIPS</b>	Regional Institute for Population Studies
<b>ROSC</b>	Report on the Observance of Standards and Codes
<b>RRSF</b>	Regional Reference Strategic Framework
<b>SADA</b>	South African Data Archive
<b>SALDRU</b>	Southern Africa Labour and Development Research Unit
<b>SCS</b>	Service Colonial de Statistiques
<b>SDA</b>	Survey Data Archive
<b>SDPA</b>	Statistical Development Programme for Africa
<b>SDDS</b>	Special Data Dissemination Standard
<b>SEAC</b>	Social and Economic Archives Committee
<b>SIGSOC</b>	Special Interest Group on Social Science Computing
<b>StatsSA</b>	Statistics South Africa
<b>StatCan</b>	Statistics Canada
<b>STATNET</b>	Statistics Technical Advisory Panel and Network
<b>STC</b>	Statistical Training Centre
<b>STPA</b>	Statistical Training Programme for Africa
<b>TAP</b>	Technical Advisory Programme, Technical Assistance Programme
<b>UBOS</b>	Uganda Bureau of Statistics
<b>UKDA</b>	UK Data Archive
<b>UKZ-N</b>	University of KwaZulu-Natal

<b>UNDP</b>	United Nations Development Programme
<b>UNECA</b>	United Nations Economic Commission for Africa
<b>UNECE</b>	United Nations Economic Commission for Europe
<b>UNESCO</b>	United Nations Educational, Scientific and Cultural Organisation
<b>UNFPA</b>	United Nations Fund for Population Activities, United Nations Population Fund
<b>UNSC</b>	United Nations Statistical Commission
<b>UNSD</b>	United Nations Statistics Division
<b>UNSTATS</b>	United Nations Statistical Commission
<b>WAEMU</b>	West African Economic and Monetary Union
<b>WAPOR</b>	World Association for Public Opinion Research
<b>WSIS</b>	World Summit on the Information Society

University of Cape Town

## GLOSSARY

### **Anonymised Microdata**

Anonymised microdata constitute “individual statistical records which have been modified in order to minimise, in accordance with current best practice, the risk of identification of the statistical units to which they relate” (King, 2003:100).

### **Confidential Data**

Confidential data is that which “allow[s] statistical units to be identified, either directly or indirectly, thereby disclosing individual information” (King, 2003:98).

### **Data**

Data is “the recorded empirical observations on the cases under study” (Lewis-Beck, Bryman & Liao, 2004:234). Data is the information we acquire on the subjects under study, and survey data can be defined as the characteristics of the cases studied in survey research. Throughout this study the term “data” will be referred to in the singular form. This is justified on the grounds that although the original Latin term would have been the plural of the word “datum”, and therefore the use of the term in the singular form would be against the rules of Latin grammar, nowadays the term is also treated as a mass noun, like the concept “information” in non-US English-speaking countries (Compact Oxford English dictionary, 2009).

### **Data Curation**

Throughout this study the term “data curation” will be used refer to digital curation of survey microdata to differentiate this from the curation of other digital information, such as administrative records. The terms “survey data curation” and “survey data management” will be used to refer to the execution of procedures related to managing the use of survey data or survey statistics.

### **Data Repurposing**

Data repurposing is “the use of data for purposes other than the original research project” (Tanenbaum & Taylor, 1991:229).

## **Digital Curation**

Digital curation can be defined as “maintaining and adding value to a trusted body of digital information for current and future use” (Digital Curation Centre, 2005).

## **Metadata**

“Statistical metadata is anything that you need to know to make proper and correct use of the real data, in terms of capturing, reading, processing, interpreting, analysing and presenting the information” (Musgrave, 2003:2). Metadata includes a detailed list of the attributes of each question asked in a survey, such as the question text and interviewer instructions (Lewis-Beck, Bryman & Liao, 2004:239). It can be viewed as “structured information that describes the dataset and its creation” (Economic and Social Data Service, 2006).

## **Microdata**

This is “aggregated numerical information relating to demographic, economic, financial, environmental, social or similar matters, at national, provincial or local level, which is compiled and analysed according to relevant scientific and statistical methodology” (South Africa. Parliament, 1999: definition xvii)

## **Re-analysis**

“A re-analysis studies the same problem as that investigated by the initial investigator; the same data base as that used by the initial investigator may or may not be used. If different, independently collected data are used to study the same problem, the reanalysis is called a replication. If the same data are used, the reanalysis is called a verification” (Fienberg, Martin & Straf, 1985:9).

## **Secondary Analysis**

“In secondary analysis, data collected to study one set of problems are used to study a different problem” (Fienberg, Martin & Straf, 1985:9).

## **Survey**

A survey is a method for collecting and analysing data. Surveys in the social sciences are concerned with the analysis of social phenomena through the collection of relevant social data, that is, data concerned with human society and its structures. What distinguishes surveys from other forms of research is the form of the data collected – which is a variable by case data grid, and the method of analysis of this data – which is concerned with describing and

explaining the characteristics of cases (these cases can be people or they can be other units of analysis, such as countries) (Lewis-Beck, Bryman & Liao, 2004:1102-3; de Vaus, 2002:3-4). The two relevant survey types referred to in this research are the census (which attempts to collect data from all members of a population) and the sample survey (which collects data from a sample of members of a population) (Statistics Canada, 2003:7).

### **Survey Data Archive**

The International Federation of Survey Data Organisations (IFDO) defines Survey Data Archives, or as they are referred to on the IFDO website, Survey Data Organisations, in terms of their function: they are organisations that archive and disseminate data and promote and participate in the exchange of data and archival expertise regionally and internationally (Mochmann, 2005:1). These organisations “are resource centres that acquire, store and disseminate digital data for secondary analysis for both research and teaching. Their prime function is to ensure long-term preservation and future usability of the data they hold” (Lewis-Beck, Bryman & Liao, 2004:234). The function of these organisations is also to facilitate data sharing: that is to ensure that data from social surveys is made available to researchers for re-use, by acquiring, processing and storing this data and distributing it to the academic community for secondary analysis (Mochmann, 2005:1).

The organisation under scrutiny in this investigation is the Survey Data Organisation as described above. However, the term Survey Data Archive (SDA) will be used for a number of reasons. Firstly, although the term “Survey Data Organisation” is used by IFDO to refer to these institutions, the term Survey Data Archive is internationally accepted and is the term IFDO members use to refer to themselves (International Federation of Data Organisations for the Social Science [IFDO], 2005).

Secondly, although the chosen terminology may imply these organisations only create and manage data archives this is no longer the accepted meaning of the term. They disseminate data, and they play an important role in promoting survey research and setting and maintaining data standards, nationally and internationally. This includes fulfilling a quality-control role with regard to national datasets. Thus the term “archive” could imply a narrower role than these organisations fulfil. However, the role of archives has changed over the years, and the term has come to be associated with the dissemination and promotion of collected data, and not merely its curatorship.

Thirdly, the emphasis in this study is on organisations that archive and disseminate census and survey data, and not information from social research in general. The IFDO definition of data archives does not specifically refer to the data in question as being census or survey information, that is, the data files and documentation resulting from censuses and survey research. The assumption can be made that the data referred to relates to censuses and surveys, however, as census and survey datasets constitute the predominant holdings of IFDO affiliates (Mochmann & de Guchteneire, 2005). However, to ensure clarity regarding the concept, the term used in this study includes the word “survey”. This term will be used for the purpose of distinguishing SDAs from archives of social science data working with sources other than census and survey microdata.

A relevant question is whether SDAs as a category should include the official state archives of a country. A SDA refers specifically to institutions that collect and disseminate survey data. This type of data should be the focus of their collection for them to be referred to by this term. National or State Archives manage and preserve public records, in many forms (South Africa. National Archives and Records Service, [2009?]). Their holdings comprise the historical records of government activities, in many formats. Therefore, national archives do not fall under the definition of SDAs used in this study.

This definition of SDAs does not encompass the National Statistics Office (NSO) responsible for conducting censuses and surveys in each country. NSOs are responsible for the collection and distribution of national statistics, but not necessarily in the form of microdata. However, in some Eastern European countries, and in Africa, the official statistical bodies are the main, sometimes the only, organisations archiving social survey data (Hausstein & de Guchteneire, 2002:69). For this reason it would be remiss to exclude these official agencies from any study of SDAs. However, as they are not strictly SDAs they will be examined as a separate category of data producer and supplier in countries in Africa where they perform this function.

An important aspect of the definition of this type of institution is whether their charging for dataset usage should disqualify them as SDAs for the purpose of this research. Charging for services provided has long been an issue in the management of SDAs. While many SDAs provide datasets and their documentation free of charge, or at a nominal cost, some charge for data usage or have institutional membership. Costs of access to data may be nominal or quite substantial. National data archives in Europe are state-funded. Archives attached to academic institutions in the US and Canada derived a certain percentage of their funding from usage or

membership fees. This is regarded as necessary to supplement their other funding sources, such as government grants, private sector research funding or financial inputs from their academic institutions or departments (Clubb, et al., 1985:65). However, open access does not necessarily mean free access (Fienberg, Martin & Straf, 1985:76). For this reason those organisations charging membership fees, and those charging for usage of certain datasets in their collections, still fall under this definition of SDAs.

### **Survey Data Archiving**

The archiving of survey data involves the storage and dissemination of data from survey research, and its documentation (Lewis-Beck, Bryman & Liao, 2004:235; Munoz, 2005:306). Data archiving may also incorporate the addition of value-added input to the final survey dataset made available to researchers, such as additional information on the data files, or suggestions on appropriate usage of the data. Contemporary data archiving often includes the promotion and provision of skills for secondary analysis of the survey data.

It is important for this study that a distinction is made between data management and data archiving. The latter is a subset of the former, and it is essential that data management procedures (including archiving) be part of the initial survey design. Data archiving should be part of the data management plan of survey organisations (Humphrey, et al., 2000:1). Data management begins with aspects of questionnaire design undertaken in survey research (Munoz, 2005:306) and is part of the fieldwork (in the form of data entry checks and quality control). It incorporates the data entry process, as well as the storage and dissemination (archiving) of the survey product (Munoz, 2005:333).

# TABLE OF CONTENTS

PLAGIARISM DECLARATION .....	I
ACKNOWLEDGMENTS .....	II
ABSTRACT .....	III
LIST OF ACRONYMS.....	IV
GLOSSARY .....	IX
TABLE OF CONTENTS .....	XIV
<b>PART ONE: SURVEY DATA ARCHIVES AND DATA CURATION WORLDWIDE .....</b>	<b>1</b>
1.1 AIM OF THE STUDY.....	2
1.2 BACKGROUND TO THE STUDY.....	3
1.3 RESEARCH QUESTIONS.....	4
1.3 RESEARCH METHOD.....	4
<b>CHAPTER 2: AN HISTORICAL OVERVIEW OF SDAS .....</b>	<b>7</b>
2.1 EARLY DATA REPOSITORIES .....	7
2.2 THE ESTABLISHMENT OF NATIONAL SDAS AND INITIAL CONCERNS.....	9
2.3 INTERNATIONAL COOPERATION AMONG SDAS .....	11
2.3.1 <i>The Establishment of Coordinating Bodies</i> .....	11
2.3.2 <i>The Development of Standards</i> .....	12
2.3.3 <i>SDAs and Comparative International Surveys</i> .....	13
2.4 THE TRAINING ROLE OF SDAS .....	13
2.5 PROFESSIONAL ASSOCIATIONS .....	14
2.6 NEW ROLES FOR SDAS .....	15
<b>CHAPTER 3: THE VALUE OF DATA SHARING AND THE ROLE OF SDAS IN THIS REGARD.....</b>	<b>16</b>
3.1 THE VALUE OF SURVEY DATA.....	16
3.2 THE VALUE OF DATA SHARING .....	16
3.2.1 <i>The Value of Data Sharing for Academic Research</i> .....	18
3.2.2 <i>The Educational Value of Data Sharing</i> .....	19
3.2.3 <i>The Value of Data Sharing for Economic Growth</i> .....	19
3.2.4 <i>The Value of Data Sharing for Good Governance</i> .....	20
3.3 DATA ARCHIVES AND DATA SHARING.....	21
<b>CHAPTER 4: THE ROLE OF SDAS IN SAFEGUARDING DATA QUALITY.....</b>	<b>25</b>
4.1 DEFINITION OF DATA QUALITY AND QUALITY DIMENSIONS.....	25

4.2 LEGISLATION AND GUIDELINES RELATING TO DATA QUALITY .....	26
4.3 THE ROLE OF SDAs IN PROMOTING DATA QUALITY .....	27
4.3.1 <i>Ensuring Data Relevance</i> .....	27
4.3.2 <i>Ensuring Data Accessibility</i> .....	28
4.3.2.1 <i>Countering the Dissuading Influence of the Academic Reward System</i> .....	28
4.3.2.2 <i>Facilitating Data Discovery</i> .....	30
4.3.2.3 <i>Supporting Technical Compatibility</i> .....	31
4.3.2.4 <i>The Data Quality/Data Protection Compromise</i> .....	31
4.3.3 <i>Promoting Accurate Survey Data</i> .....	35
4.3.3.1 <i>Supporting Sound Research Methods</i> .....	35
4.3.3.2 <i>Error Reporting and Correction</i> .....	36
4.3.4 <i>Promoting Timely Data Release</i> .....	37
4.3.5 <i>Supporting Data Coherence</i> .....	39
4.3.6 <i>Enhancing Data Interpretability</i> .....	40
4.3.6.1 <i>Providing Value Added Input to Documentation</i> .....	40
4.3.6.2 <i>Preventing Archiving Errors</i> .....	42
<b>PART TWO: SURVEY DATA CURATION IN AFRICA.....</b>	<b>43</b>
<b>CHAPTER 5: A SURVEY DATA CURATION HISTORY OF AFRICA .....</b>	<b>45</b>
5.1 THE COLONIAL PERIOD .....	46
5.1.1 <i>International Support for African Data Curation</i> .....	47
5.1.2 <i>African Data Curating Institutions in the Colonial Period</i> .....	47
5.2 THE 1960S.....	48
5.2.1 <i>Stakeholders in Coordination for Data Development in Africa</i> .....	49
5.2.2 <i>Improving Institutional Infrastructures to Advance Data Quality</i> .....	49
5.2.3 <i>Support for Technological Infrastructures in Africa</i> .....	50
5.2.4 <i>Human Resource Development</i> .....	50
5.3 THE 1970S AND 1980S .....	51
5.3.1 <i>Stakeholder Coordination</i> .....	51
5.3.2 <i>Support for Institutional Infrastructures to Advance the Quality of African Data</i> .....	52
5.3.3 <i>The Development of African Technological Infrastructures for Data Curation</i> .....	52
5.3.4 <i>Human Resource Development for Data Curation - 1970s and 1980s</i> .....	53
5.4 DEVELOPMENTS IN AFRICAN DATA CURATION IN THE 1990S .....	54
5.4.1 <i>Stakeholder Coordination in the 1990s: The Addis Ababa Plan of Action for Statistical Development (AAPA)</i> .....	54
5.4.2 <i>Development of African Institutional Infrastructures in the 1990s</i> .....	56
5.4.2.1 <i>National Data Producers and Support for Data Quality</i> .....	56
5.4.2.2 <i>Other Data Curation Institutions in Africa in the 1990s</i> .....	58
5.4.3 <i>ICT Infrastructures for Knowledge Management in Africa in the 1990s</i> .....	58
5.4.4 <i>Human Resource Development for Data Curation in the 1990s</i> .....	59
5.4.5 <i>Professional Associations</i> .....	60

5.5 DATA CURATION IN AFRICA FROM 2000 .....	60
5.5.1 Coordination of International Support for Data Curation in Africa from 2000 .....	61
5.5.2 Regional Initiatives .....	64
5.5.3 Development of Regulatory Infrastructures for Data Curation in Africa: 2000 Onwards .....	65
5.5.4 Institutional Infrastructures: 2000 Onwards.....	66
5.5.4.1 National Statistics Offices and Data Quality Initiatives from 2000 .....	66
5.5.4.2 Survey Data Archives in Africa .....	68
5.5.5 Support for African Technological Infrastructures for Data Curation in the 2000s .....	69
5.5.6 Human Resource Development for Data Curation from 2000.....	70
5.5.7 Professional Associations in the Region: the Establishment of the African Association of Statistical Data Archivists .....	71
5.5.8 The 2000s: The Sharing of Publicly Funded Research Data.....	72
<b>CHAPTER 6: DATA UTILISATION FOR GOVERNANCE IN AFRICA .....</b>	<b>77</b>
6.1 THE ROLE OF STATISTICS OFFICES IN NATIONAL PLANNING .....	77
6.2 EVIDENCE FOR DATA UTILISATION IN AFRICA .....	78
6.3 THE GRADUAL INFLUENCE OF RESEARCH.....	78
6.4 THE INFLUENCE OF THE POLITICAL CLIMATE ON DATA UTILISATION .....	79
6.5 THE NEED FOR CONSENSUS TO AFFECT DATA UTILISATION .....	80
6.6 ADVOCACY FOR DATA UTILISATION IN AFRICA .....	81
6.7 SKILLS LEVELS IN GOVERNMENT AND EVIDENCE-BASED POLICY-MAKING .....	83
6.8 THE ROLE OF THE RESEARCH-POLICY INTERFACE IN DATA UTILISATION .....	83
6.8.1 Historical Antipathies .....	84
6.8.2 The Changing Research-Policy Interface in Africa .....	86
6.9 DATA QUALITY AND EVIDENCE-BASED DECISION-MAKING IN AFRICA: THE ROLE OF SDAs .....	88
<b>CHAPTER 7: OBSTACLES TO EFFECTIVE SURVEY DATA CURATION IN AFRICA.....</b>	<b>91</b>
7.1 INFRASTRUCTURAL OBSTACLES TO DATA CURATION IN AFRICA .....	91
7.1.1 Problems with Regulatory Infrastructures .....	92
7.1.1.1 Limited Focus on Data Curation in Existing Information Policies .....	92
7.1.1.2 Lack of Integration of Information Policies with Development Agendas .....	94
7.1.1.3 Donor Driven Agendas .....	95
7.1.1.4 Minimal Private Sector Support.....	95
7.1.1.5 Confidentiality Concerns in Policymaking .....	96
7.1.1.6 Other Hindrances to Effective Data Curation Policies.....	97
7.1.2 Problems with Institutional Infrastructures .....	97
7.1.2.1 African National Statistics Offices .....	97
7.1.2.2 Absence of Other Data Curating Institutions .....	102
7.1.3 Poor Technological Infrastructures .....	103
7.2 HUMAN RESOURCE CONSTRAINTS TO DATA CURATION ON THE CONTINENT .....	106
7.2.1 Shortage of Skills to Support ICT Infrastructures for Data Sharing .....	106

7.2.2 Inadequate Statistical Training.....	107
7.2.3 A Paucity of Data Management Skills.....	108
7.2.4 Limited Research Capabilities.....	109
7.3 POOR CO-ORDINATION AMONG STAKEHOLDERS IN DATA EXCHANGE IN AFRICA.....	109
7.3.1 Problems with International and Regional Coordination.....	110
7.3.2 Problems with National Coordination.....	110
<b>CHAPTER 8: CONCLUSION AND RECOMMENDATIONS .....</b>	<b>113</b>
8.1 CONCLUSION: POLITICAL FACILITATION FOR DATA CURATION IN AFRICA.....	113
8.2 RECOMMENDATIONS REGARDING GOVERNMENTAL SUPPORT FOR DATA CURATION.....	114
8.2.1 Policies for Data Curation Advocacy.....	114
8.2.2 Policies for Funding of Data Curation.....	115
8.2.3 Counteracting Bureaucratic Constraints.....	116
8.2.4 Policies to Advance Local Capacities for Data Curation.....	116
8.2.5 Support for Professional Associations.....	117
8.2.6 Establishment of the Regulatory Infrastructure for Data Curation.....	117
8.2.7 Development of the Necessary Technological Infrastructure.....	118
8.2.8 Development of Institutional Infrastructures for Data Curation.....	118
8.2.8.1 SDAs within NSOs.....	119
8.2.8.2 SDAs within African Universities.....	120
8.2.8.3 SDAs as Autonomous Government Institutions.....	121
<b>REFERENCES.....</b>	<b>124</b>
<b>APPENDICES.....</b>	<b>158</b>
APPENDIX A: TIMELINE OF SURVEY DATA CURATION WORLDWIDE.....	159
APPENDIX B: SURVEY DATA ARCHIVES WORLDWIDE.....	172
APPENDIX C: AFRICAN DATA CURATION TIMELINE.....	177
APPENDIX D: AFRICAN MICRODATA ACCESS SURVEY 2009.....	184
APPENDIX E: AFRICAN MICRODATA CURATION INDEX.....	193

## LIST OF PHOTOGRAPHS

PHOTOGRAPH 1: CENTRO DE INVESTIGACIONES SOCIOLOGICAS (CIS), THE NATIONAL SDA IN SPAIN (PHOTO: CIS WEBSITE, 2009: HTTP://WWW.CIS.ES).....	6
PHOTOGRAPH 2: THE INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH (ICPSR) IN THE US (PHOTO: SHAW LACY: PANORAMIO WEBSITE, 2009: HTTP://WWW.PANORAMIO.COM/PHOTOS/ORIGINAL/5201386).....	6
PHOTOGRAPH 3: THE NORWEGIAN SOCIAL SCIENCE DATA SERVICES (NSSDS) (PHOTO: NSSDS WEBSITE, 2008: HTTP://WWW.NSD.UIB.NO/COMMON/IMG/ILL_NSDHUS_LARGE.JPG).....	24
PHOTOGRAPH 4: THE DATA ARCHIVE OF THE INSTITUTE OF SOCIOLOGY, THE SLOVAK REPUBLIC (SASD). (PHOTO: SASD WEBSITE, 2008: HTTP://WWW.SOCIOLOGIA.SAV.SK).....	24
PHOTOGRAPH 5: THE NATIONAL INSTITUTE OF STATISTICS AND ECONOMIC ANALYSIS (INSEA) IN BENIN (PHOTO: INSEA WEBSITE, 2008: HTTP://WWW.INSAE-BJ.ORG).....	44
PHOTOGRAPH 6: THE INSTITUTE OF NATIONAL STATISTICS AND ECONOMIC AND DEMOGRAPHIC STUDIES (INSEED) IN CHAD. (PHOTO: INSEAD WEBSITE, 2009: HTTP://WWW.INSEED-TCHAD.ORG).....	44
PHOTOGRAPH 7: THE CENTRAL AGENCY FOR PUBLIC MOBILIZATION AND STATISTICS (CAPMAS) IN EGYPT (PHOTO: CAPMAS WEBSITE, 2008: HTTP://WWW.CAPMAS.GOV.EG).....	76
PHOTOGRAPH 8: THE GAMBIA BUREAU OF STATISTICS (GBOS). (PHOTO: GBOS WEBSITE, 2009: HTTP://WWW.GBOS.GM).....	76
PHOTOGRAPH 9: THE LIBERIA INSTITUTE OF STATISTICS AND GEO-INFORMATION SERVICES (LISGIS). (PHOTO: LISGIS WEBSITE, 2009: HTTP://WWW.LISGIS.ORG).....	90
PHOTOGRAPH 10: THE PUBLICATIONS DIVISION OF THE NATIONAL STATISTICAL OFFICE (NSO) OF MALAWI. (PHOTO: MALAWI NSO WEBSITE, 2009: HTTP://WWW.NSO.MALAWI.NET).....	90
PHOTOGRAPH 11: THE LESOTHO BUREAU OF STATISTICS (LBOS). (PHOTO: LYNN WOOLFREY, 2008).....	112
PHOTOGRAPH 12: NADA TRAINING AT THE LESOTHO BUREAU OF STATISTICS (LBOS). (PHOTO: LYNN WOOLFREY, 2008).....	112
PHOTOGRAPH 13: DATAFIRST STAFF (INCLUDING THE AUTHOR) AND WORLD BANK CONSULTANTS VISITING THE CENSUS 2007 WAREHOUSE IN MOZAMBIQUE IN 2008. (PHOTO: GEOFFREY GREENWELL, 2008).....	123
PHOTOGRAPH 14: PROCESSING THE 2007 MOZAMBIKAN CENSUS RESULTS. (PHOTO: GEOFFREY GREENWELL, 2008).....	123

## LIST OF MAPS

MAP 1: AFRICAN MICRODATA ACCESS SURVEY 2009.....	101
--	-----

**PART ONE: SURVEY DATA ARCHIVES AND DATA CURATION  
WORLDWIDE**

## CHAPTER 1: INTRODUCTION

### 1.1 Aim of the Study

Of concern to this study is what constitutes the most effective way of managing the information garnered from social surveys in Africa to facilitate its re-use. The study will address the why, where and how questions around survey data archiving<sup>1</sup> on the continent.

Utilising survey data for the study of social phenomena other than those for which the original survey was initiated is a relatively new research approach in Africa. Thus, best practice for this type of research is still being put in place by African organisations. Social survey information refers to both the statistical data which is the final product of censuses or sample surveys, and the documentation provided with the data to facilitate its reuse. Documentation includes technical notes and questionnaires from the survey process, as well as metadata (detailed information about the data) and reports produced to add value to the final survey findings. The optimal management of this type of information to facilitate its reuse is the subject of this research.

Part one of this study will investigate data management practices worldwide to provide background to the research. This will begin with chapter two which will be an historical overview of the preservation and usage of social survey data internationally, in the context of the development of Survey Data Archives (SDAs) as data curating organisations. An historical timeline of developments in this regard will be included as Appendix A. Appendix B will provide a comprehensive list of SDAs worldwide for additional information.

Chapter three will present a case for the curation of social survey data as a national and regional resource for scientific investigation and sound national decision-making and will suggest SDAs as appropriate organisations for the storage and dissemination of this information. Chapter four will address data management best practice within these organisations to promote data quality, that is, to assure the relevance, availability, and usability of survey datasets for research purposes.

---

<sup>1</sup> For a definition of this and other terms needing clarification, see the Glossary.

## INTRODUCTION

Part two of this research will focus on the curation of survey data in the African context. Chapter five will constitute a descriptive overview of the current state of survey data curation in Africa. Chapter six will provide an historical account of data management in African countries. This will be supported by an African data curation timeline supplied as Appendix C. Chapter seven will examine survey data usage for policymaking by African governments, and investigate those factors that foster evidence-based decision-making. Chapter eight will elaborate on the obstacles to effective data curation in the region, including current impediments to the regional sharing of government survey microdata. The results of an investigation of access to African microdata will be included as Appendix D to provide empirical data in this regard. An African microdata curation index (Appendix E) will provide a summary of the role of NSOs in data curation on the African continent.

Chapter nine will present a conclusion and recommendations to foster the effective curation of national survey data as a development resource in African countries. Strategies will be suggested for the establishment of institutions to promote data management and data re-use in the region. SDAs will be examined in this regard and the study will elaborate on the possibilities of and procedures for the creation of a Pan African SDA network. The European model will be used as a basis for this, and adaptations to this will be discussed with regard to local relevance.

### **1.2 Background to the Study**

Social surveys are an important source of data for the analysis of social conditions in Africa for effective development planning. The data from these studies provides vital information on the state of nations. Governments around the world have come to appreciate the value of survey data and the relevance of studies based on secondary analyses of this data for policy formulation. As a result demands are being made by policymakers for survey products that are relevant and current. Academic institutions and the private sector have also come to expect optimum usability from survey data for repurposing, particularly in terms of the consistency and comparability of survey datasets. How, then, do survey teams ensure that the final product of their research is accessible and usable? This research addresses these issues by looking at historical developments and current practices in the data archiving aspects of survey research worldwide and with reference to the situation in Africa, where suggestions are made for improvements in this regard.

## INTRODUCTION

### 1.3 Research Questions

This study will investigate the purpose of curating social survey microdata, and ascertain the value of such collections. Current data management and data dissemination practices in African countries will then be examined, particularly with regard to official African data. Obstacles to the effective curation of African microdata will be elaborated on and strategies to overcome these impediments will be suggested. Questions dealing with this issue will be concerned with the appropriate mechanisms that need to be put in place to affect long-term data preservation and data exchange in African countries to foster the reuse of African survey statistics to support effective policymaking. Questions to be addressed in the thesis include:

- What purpose does the collection of survey and census microdata serve with regard to supporting government planning and academic research?
- Do African governments and academics make use of this type of information and what impediments exist with regard to effective data usage?
- What institutional support exists to assist with survey and census data usage for these purposes?
- Are these structures available in African countries for this purpose, or is such research support restricted to developed countries?
- What institutional structures are appropriate for fostering data usage in African countries, and how can these be established and maintained?

### 1.3 Research Method

A literature study was undertaken to address questions around the purpose of collecting and managing survey and census data for re-use. Literature examined included documents related to the establishment of survey data archives in Europe and North America, including documents of international donor organisations supporting data curation in these regions. Documentation from the early years of the establishment of Survey Data Archives were also examined to provide material on the increasingly important role these organisations play in assisting data usage by academics and government officials.

To develop an understanding of the state of survey data curation in Africa, an extensive review of the websites of National Statistics Offices in all fifty-three African countries was undertaken. Documentation produced by these institutions related to data management was also included in the review. Seminal documents related to support for the usage of official data were

## INTRODUCTION

examined, including government documents and publications from key meetings initiated by international donor organisations. The relevance of this methodology for this purpose is that it provided the researcher with a thorough overview of historical developments in the field, which revealed data curation trends in the region, as well as recurring obstacles to effective data usage in African countries. The limitations of this approach include the bias with regard to the view of international donor organisations, which produce the majority of publications dealing with data curation in Africa, from a particular standpoint. Very little documentation has been produced in African countries on the topic and this could contribute to a less well-rounded view of the data curation situation in Africa.

An empirical survey of access to microdata from African governments was chosen as the method adopted to examine these obstacles in more detail. The National Statistics Offices of all fifty-three African countries were contacted in this regard, using details obtained from their websites, and an attempt was made to obtain survey microdata from these sources. This further elaborated on impediments to data management for planning and research in the region. The limitations of this method was that technological constraints prevented communication with many African National Statistics Offices with the result that these less well-resourced institutions were not able to provide input on the topic. An attempt was made to counteract this limitation by soliciting information from representatives of these institutions as conferences and workshops in an attempt to obtain accurate information on the situation with regard to survey data management in their countries.

## INTRODUCTION



Photograph 1: Centro de Investigaciones Sociológicas (CIS), the national SDA in Spain (Photo: CIS website, 2009: <http://www.cis.es>).



Photograph 2: The Inter-University Consortium for Political and Social Research (ICPSR) in the US (Photo: Shaw Lacy: Panoramio website, 2009: <http://www.panoramio.com/photos/original/5201386>).

## CHAPTER 2: AN HISTORICAL OVERVIEW OF SDAS

*Data management does not terminate with the publication of the first statistical reports” (Munoz, 2005: 306).*

This chapter examines the history and development of SDAs as repositories of survey microdata and as data sharing facilities. An historical timeline of international developments with regard to these institutions is included as Appendix A to provide further information for this chapter. A list of current SDAs, which includes information on their institutional locations, web addresses and organisational affiliations, is provided as Appendix B.

### 2.1 Early Data Repositories

The establishment of dedicated archives for the sharing of social survey data came about as a result of the development of quantitative research within the social sciences involving the collection of social statistics through censuses and sample surveys. Survey research was seen to aid the examination of societies in order to bring about positive changes in these societies. Quantitative social research and the establishment of data archives to house and disseminate the output of this research were in turn aided by the invention of the computer and its increasingly sophisticated technologies (Rowe, 1999:4).

Survey research began in the 1930s and developed rapidly, as this form of investigation provided researchers and policymakers with valuable data on social conditions (Kalton, 2000:3). Market research surveys also yielded valuable data on target populations. The many market research surveys conducted in the 1950s produced large amounts of machine-readable data stored on punched cards, and the availability of this data led to its repurposing by researchers as they realised the data could be a useful research resource (Tanenbaum & Taylor, 1991:229). For example, data from market research surveys of a particular target group could be used to investigate social phenomena in the same group. This stimulated a need for facilities to preserve the original data files from surveys and make them available in a format suitable for secondary investigation. The computer revolution assisted this, as it led to faster and more efficient data processing and analyzing. The evolution of sophisticated and user-friendly statistical analysis packages allowed researchers to conduct in-depth analyses on survey data, and in some countries introduced even university undergraduates to secondary analysis of data

## HISTORICAL OVERVIEW

(Scheuch, 2003:389). This in turn led to increased demands from data users for more accurate and timely data which served to advance survey methodology (Kalton, 2000:5).

SDAs began as storage facilities for survey datasets, and archive staff gradually took on a more active role in disseminating this data and providing value-added services to data users (Tanenbaum & Taylor, 1991:225). Their use of advances in computer technology meant they were able to optimise data access, including converting data for use across computer systems. For example, staff at the UK Data Archive developed conversion software to make official data from government surveys available in a format accessible to researchers (Tanenbaum & Taylor, 1991:227).

In 1946 the first SDA, The Roper Public Opinion Research Center, was established at Williams College, Massachusetts in the US (Rowe, 1999:5). It was set up by Elmo Roper to house the results of market research surveys conducted by him from the 1930s (Roper Center, 2009; Scheuch, 2003:385). This archive was opened to the public in 1957 (Scheuch, 1990:95). The founders of the Roper Centre established the standards for similar organisations when they stated that the function of an archive was the collection and preservation of survey data from research organisations in order to make this data available for secondary analysis and for teaching purposes (Roper Public Opinion Research Center, 1963). Eventually, data archives and their clients began to have a reciprocal relationship, as academic research fed into new surveys, and this survey information was deposited in the archives. Social surveys conducted by government statistics offices, academic institutions and private research organisations eventually replaced market surveys as the main holdings of these archives (Tanenbaum & Taylor, 1991:226).

International support for data sharing was evidenced as early as 1950, when UNESCO set up an International Committee for Social Sciences Documentation (ICSSD), tasked with international coordination of social science research (The International Committee for Social Sciences Documentation, 1962:178). The ICSSD was a specialist body of the UN's International Social Science Council (ISSC) and its specific activities were concerned with creating access to social science information and formulating and promoting the use of methods for presenting reference works in this field. This was done in those early days through the creation of bibliographies and technical documents. One of the mandates of the ICSSD was to examine procedures which could be used to create "National Documentation Centres" for the social sciences (The International Committee for Social Sciences Documentation, 1962:187). A leading figure in the archive

## HISTORICAL OVERVIEW

movement, political scientist Stein Rokkan, based at the University of Bergen, became the chairperson of the ISSC's Standing Committee on Social Science Data Archives when this body was established, and documented the need for repositories of survey data to aid social research (Rowe, 1976:97; Scheuch, 2003:386,391).

### **2.2 The Establishment of National SDAs and Initial Concerns**

The 1960s saw the establishment of several European SDAs. The earliest European archive, the Zentralarchiv in Cologne, Germany, was established in 1960, initially as an institute of the University of Cologne. Local archives were constituted in the Netherlands and Norway in the same year. In 1962 twenty four universities in the United States formed the Inter-University Consortium for Political and Social Research (ICPSR) to facilitate access to survey data for the teaching of university courses. The ICPSR acted as an archive for machine-readable data for its member institutions. Other US data archives were established in the 1960s, such as the data facility at the University of California in Berkeley, founded in 1964 (Scheuch, 1990:97).

Debates around the role and structure of SDAs soon emerged among the new archives. One such debate concerned the value of developing an international SDA as a central location for all survey products, as opposed to the development of decentralized national collections. Another controversial issue concerned the funding of archives. The question was raised whether SDAs should be self-funded organisations, recouping costs from membership fees and charging for services, or receive state funding. Legislative and ethical issues around survey data management were also addressed by archive administrators, such as data ownership issues. SDAs also began to develop their role as organisations promoting data quality.

The idea of a central world data archive was initiated by the Roper Center. The Center attempted to draw in worldwide data to form an international archive of data holdings (Scheuch, 1990:98). Its attempts to add European surveys to its collection met with resistance and spurred archives in Europe to organize a joint strategy to resist this endeavour. European archive directors were united in their defence of national collections, as opposed to the World Data Archive proposed by the Roper Center. Cooperative arrangements among nascent data archives in Europe were initiated in 1962 at the first UN conference on Social Science Data Archives in La Napoule, France, organized by the UN International Social Science Council (ISSC). A year later, at a conference of the World Association for Public Opinion Research (WAPOR), delegates were critical of the concept of a single international SDA based in North America.

## HISTORICAL OVERVIEW

Szalai and Petrella noted in 1977 that US participants were accused of practicing “data imperialism” and conducting “safari type of research” by “going into another country, paying the ‘natives’ for collecting responses, then taking the data back home” (Scheuch, 2003:387).

Such emotional responses to suggestions of a world data archive reflected participants’ fears of loss of control over national data. The controversy raised the issue of data ownership among SDAs. The European delegates feared loss of access to their own data once it became part of a central collection (Scheuch, 1990:98). Issues of control over data quality were also raised in support of decentralized archives (Scheuch, 2003:388). International co-operation between national data archives was presented as an alternative to a centralized collection (Scheuch, 1990:98).

The differing views of the US and European SDA Directors included those around the funding of data archives. From their establishment, US and European archives had different positions on charging for services. Traditionally, public services in Europe, such as archives, have been subsidized. This has never been the case in the United States, where historically social science service facilities have received scant public funding (Scheuch, 1990:98-101). These different traditions are reflected in the fact that most US SDAs expect some form of payment from clients using the data. For example, the ICPSR obtains subscriptions from institutional members, who are then given rights to download datasets (Inter-University Consortium for Political and Social Research, 2009). The Roper Center, from its inception, charged for postage and materials delivered to clients, and for labour costs (Roper Public Opinion Research Center, 1963:120). European archive directors, however, were unimpressed with what they perceived as the commoditisation of survey data, whereas, as publicly funded organisations, they were in a position to treat their data holdings as a resource for all to use.

The second ISSC Conference on Data Archives was held in Paris in September 1964. At this conference European institutions formally rejected the notion of a world data archive based in the US and the notion of a federation of data archives was agreed upon, with the European archives securing government funding support. Delegates also agreed that the focus of these archives should be on survey data, as opposed to other research findings (Scheuch, 2003:388-389).

New SDAs continued to be established during this period. In the United Kingdom a data archive was set up in 1967 at the University of Essex, with a mandate to provide training in data

## HISTORICAL OVERVIEW

management to all UK universities. The support infrastructure for data archives during this period came from UNESCO's International Social Science Council which formed a Standing Committee on Social Science Data Archives (CSSDA) (Mochman, 2005:1). The establishment of these committees resulted from the work of academics such as Stein Rokkan, and the sociologist Erwin Scheuch, of the University of Cologne. These researchers played a pivotal role in lobbying the ISSC for the necessary infrastructure to promote their comparative quantitative social research. Erwin Scheuch was appointed as the Chairman of the CSSDA from 1966. Rokkan was president of the International Social Science Council from 1973-77, during which time he promoted the development of the European SDA network (Scheuch, 2003:385,391).

### **2.3 International Cooperation among SDAs**

The Helsinki Act of 1975 saw thirty five countries - the United States and Canada, the Soviet Union and the countries of Europe - sign an agreement pledging their endorsement of international economic and political cooperation. The accord included coordination of information exchange among the signatories (Stamatiou, 1988:600). This gave encouragement and international funding support to co-operative efforts at social science data sharing among European and North American archives.

#### **2.3.1 The Establishment of Coordinating Bodies**

International coordination of data resources was further boosted by the establishment of the European Cooperation in Social Science Information and Documentation (ECSSID) Project by the European Coordination Centre for Research and Documentation in Social Science (known as the Vienna Centre, a body of the UNESCO International Social Science Council). ECSSID was tasked with coordinating European efforts in developing and standardising social science documentation, and held Biennial conferences related to these issues (Rozsa, 1981:559-560).

A meeting in 1976 of the ISSC Standing Committee on Social Science Data Archives (CSSDA) saw representatives from the Roper Center, now holding the largest collection of survey data in the world, reiterate their intention to establish a world data archive in Williamsburg in the US, despite strong objections by representatives from European archives. In response to this perceived threat to their autonomy, the European archives finally took combined action and formed the Committee for European Social Science Data Archives (CESSDA), an umbrella body for regional co-operation (Scheuch, 2003:392). CESSDA was tasked with promoting the

## HISTORICAL OVERVIEW

exchange of social survey data among its membership, technical transfer among archives, and nurturing new data archives in the region. An additional role for the coordinating body was the creation of standards for describing and documenting the data holdings of member archives. A further meeting of SDAs in 1976 consolidated the principle of one data archive per country, promoted by the European archives who continued to resist contributing their holdings to a world data archive based in the US (Scheuch, 1990:99). This is how data archives developed in all countries except North America. In the US and Canada SDAs were formed in several major universities, as well as in commercial survey research organisations (Rowe, 1974-5:32-33).

In 1977 CESSDA members met with the major US archives to form the International Federation of Survey Data Organisations (IFDO). IFDO took over the role of international coordination of data archives, while CESSDA confined itself to the concerns of European archives. IFDO eventually replaced UNESCO's Standing Committee on Social Science Data (CSSD) as the international body concerned with the preservation and sharing of data from survey research (Mochmann, 2005:2). The Roper Center gave up their idea for a global archive and eventually became an IFDO affiliate (Scheuch, 2003:392).

IFDO continues to foster data curation internationally, for example by assisting in the creation of an East European Data Archive Network (EDAN) in 2001, which coordinates archive development in Eastern Europe (Haustein & de Guchteneire, 2002:18-19). CESSDA continues to take an active role with regard to data sharing projects in Europe. An important contribution to their task was a decision by the European Strategy Forum on Research Infrastructures (ESFRI) to award CESSDA a grant of 2.7 million Euros in 2008 to upgrade its research infrastructure. ESFRI was formed in 2002 to assist in the scientific integration of Europe. Seventeen European SDAs, led by the UK Data Archive, are involved in the CESSDA Project which aims to utilise software tools to provide simple, seamless access for researchers across Europe to the holdings of European SDAs and support training for data managers in the region (The CESSDA ESFRI Project, 2008; Council of European Social Science Data Archives [CESSDA], 2008).

### **2.3.2 The Development of Standards**

The first attempts at standardising study descriptions were made in the 1960s by the Steinmetz Archive in the Netherlands, the Zentralarchiv (in Germany) and the UK Social and Economic Archives Committee (SEAC) for a UK inventory of archives. Further refinement for survey documentation was sought by these archives in cooperation first with the Standing Committees of

## HISTORICAL OVERVIEW

the ISSC, and then with IFDO. Here the ICPSR in the US took the lead in developing a format for machine-readable codebooks. These codebooks are instructions for interpreting the data files, showing how information obtained is represented by numeric data, and their standardisation is vital for comparability among survey datasets (Lewis-Beck, Bryman & Liao, 2004:136; Scheuch, 2003:393).

### **2.3.3 SDAs and Comparative International Surveys**

Data managers have looked to the day when SDAs would “no longer understand themselves predominantly as the passive recipients of data deposited with them, but also create data through various merging procedures” (Scheuch, 2003:396). Historically, SDAs have only been curators and disseminators of survey data, but recently they have moved towards a more active role in promoting survey re-analysis, particularly with regard to comparative research, through the creation of comparative datasets from international surveys. For example the Zentralarchiv in Cologne is responsible for merging the datasets from the individual Eurobarometer public opinion surveys conducted regularly in several European countries by the European Commission (Scheuch, 2003:393). In this way SDAs are improving on a situation which has been commented on thus:

The creation, maintenance and dissemination of standardized micro-datasets is obviously a critical requirement for comparative research, yet this is far from being a universal feature of international survey programmes” (Verma, 2002:6).

European Data Archives have supported an initiative by the US National Science Foundation (NSF) and the UK Economic and Social Research Council (ESRC) to investigate data needs for cross-national research and to design mechanisms for the sharing of this data. This culminated in the launch of the International Data Forum (IDF) in 2007 to build international collaborative efforts for data collection and data sharing (The Foundation Conference for an International Data Forum, 2007:1).

### **2.4 The Training Role of SDAs**

Advances in technology led to the need for the development of new data curation and data analysis skills among data archivists and survey researchers. In response, SDAs took on a training role with regard to survey research and survey data management. This included instruction in survey analysis (Scheuch, 2003:390). The ICPSR initiated its first “Summer Training

## HISTORICAL OVERVIEW

Programme in Quantitative Methods of Social Research” in 1963. This programme is still being run today and includes courses on data preservation (Inter-University Consortium for Political and Social Research [ICPSR], 2009).

In 1964 the Zentralarchiv in Cologne, Germany started their spring and summer seminars which provided practical training in quantitative social and historical research. The Essex Summer School on Social Science Data Analysis and Collection began in 1968, initially funded by UNESCO. The Oslo Summer School (OSS) in Comparative Social Science Studies was launched at the University of Oslo in Norway in 1992, and offered its first courses in 1993 (Oslo Summer School in Comparative Social Science Studies, 2009). These training courses were attended by young researchers from around the world, and a major benefit of these programmes was the resultant formation of an international network of scholars interested in the re-examination of survey results for scientific advancement (Mochmann, 2005:7; Scheuch, 2003:391). This has helped create the user base necessary to ensure the continuation of SDAs’ role in research support for survey data analysts.

### **2.5 Professional Associations**

The Association of Public Data Users (APDU) was set up in the 1970s and examined data issues from the perspective of the secondary data analyst. This group investigated the possibility of publishing directories of available datasets and compared the suitability of various software programmes for analysing survey data. They attempted to persuade researchers to release their final data products for repurposing, with varying degrees of success (Rowe, 1976:99). At this stage of social survey research principal investigators were not required to release the raw data of their research even to their funding agencies, let alone place them in the public domain (Rowe, 1976:99). The final product of most social research involving surveys was a detailed report or summary document of the findings of the investigation, and there was a general lack of awareness of the value of data reuse.

In 1974 an international professional association of data managers was established, with mainly North American membership. The International Association for Social Science Information Services and Technology (IASSIST) was formed at an International Sociological Association workshop attended by SIGSOC members (the Special Interest Group on Social Science Computing of the Association for Computing Machinery (ACM)). IASSIST members were concerned with the policies and processes involved in managing data. They took part in

## HISTORICAL OVERVIEW

international training workshops and investigated confidentiality laws regarding information access. This group also examined best practice with regard to software and hardware for use in data curation and attempted to standardise metadata and file formats for survey datasets (Rowe, 1976:97-98). Issues of standardisation, comparability, legality and training are still part of this group's concerns today (International Association for Social Science Information Services and Technology, 2009).

### **2.6 New Roles for SDAs**

Historically SDAs have been concerned with building and optimizing their data collections, as well as the standardization and sharing of data resources. Working towards enhancing data standards is still a key element of archive work, especially with regard to cooperation among archives in the combining of datasets to create international resources for research. The other focus among European and North American SDAs which has become dominant is the ongoing promotion of skills, both within the survey research community, and among data archive staff (Scheuch, 2003:396). IASSIST plays a key role in organizing conferences around issues of staff training and development. SDAs are also playing a proactive role in promoting data curation best practice, for example concerning the handling of sensitive data.

Possibly inevitably, considering their intimate involvement with data and technology, SDAs have become software developers. The European Centre for Analysis in the Social Sciences (ECASS) in the Institute for Social and Economic Research (ISER) at the University of Essex (which houses the UK Data Archive) has collaborated with the Danish and Norwegian Data Archives to develop the Networked Social Science Tools and Resources (NESSTAR) web application for data management and data dissemination with funding assistance from the European Union (EU) (Networked Social Science Tools and Resources, 2009; Scheuch, 2003:397; UK Data Archive, 2007).

This chapter presented an historical overview of the development of SDAs worldwide as institutions for the preservation and management of survey microdata. Chapter three elaborates on the role of these institutions in supporting the sharing of microdata for research purposes.

## **CHAPTER 3: THE VALUE OF DATA SHARING AND THE ROLE OF SDAS IN THIS REGARD**

This chapter examines how empirical information collected through censuses and sample surveys can be a valuable resource for government planning and academic research and can be used as an effective teaching tool. The chapter stresses that effective data sharing mechanisms need to be put in place to enable the optimal usage of this type of information for these purposes.

### **3.1 The Value of Survey Data**

The literature on data archiving confirms the importance of social survey data as a research and policy tool. The value of accurate and accessible social survey data can hardly be overestimated. Government agencies have come to rely on social statistics generated by national surveys to formulate policies, particularly with regard to economic growth and poverty alleviation. There has also been a growing awareness among academics of the value of this data beyond the initial survey purpose and this new understanding among researchers and the rising costs of conducting original surveys has led to requests from academics for data producers to share their survey data (Fienberg, Martin & Straf, 1985:3-4). Advances in technology and the increasing demand for empirical data by decision-makers and academics has in turn led to an exponential growth in both survey research and the availability of survey data for secondary analysis (Kalton, 2000:3; Seekings, 2002:11). These developments, coupled with technological advances supporting the efficient storage and distribution of data, led to the establishment of dedicated SDAs as facilities for the sharing of social survey research data.

### **3.2 The Value of Data Sharing**

The findings of early surveys were published in reports and the original data files remained under the ownership of the principal investigators. Data files were sometimes stored but were not used for reanalysis or secondary analysis. Through informal academic networks, researchers began to utilise the data from surveys to investigate social issues not dealt with in the initial study. Survey data was shown to have worth beyond the original research findings and data sharing came to be valued as a cost-effective means of furthering social research (Fienberg, Martin & Straf, 1985:3-5).

## SURVEY DATA ARCHIVES AND DATA SHARING

There are no comprehensive statistics on data sharing, or on the repurposing of shared data (Boruch, 1985:113-114). However, the development of SDAs, and the growth in the size of their collections, provides some indication of data demand, and demonstrates the willingness of data producers to make their data available for further research. Another indication of the perceived value of data reuse is the funding by major donors of large-scale survey projects designed to provide research data for the wider academic community, both locally and internationally. An example of this is the General Social Survey, which has international participation and wide usage (Clubb et al., 1985:52).

It is also possible to get an idea of the demand for data by examining the statistics maintained by SDAs (Clubb et al., 1985:49-50). These include statistics of requests for data, derived from online data request forms, and records of dataset distribution. For example, the ICPSR, a major SDA in the US, indicates on its website that over the period 2001-2005, staff distributed 1906 datasets to members. This archive has a membership of more than 500 organisations, in the US and other countries, and the number of affiliates to this archive provides further indication of the research community's belief in the value of data re-use (Inter-University Consortium for Political and Social Research, [2009?]b).

The extent of actual usage of shared data is more difficult to assess. Many SDAs have web-based inventories of publications that have utilised the datasets they distribute as secondary analysts pledge to notify archives of any publications they produce using the datasets. From these can be ascertained at least the quantity of works produced with shared data. However, frequency counts cannot measure the quality of the academic output from secondary usage of data, and its impact on policy formulation (Boruch, 1985:114). This suggests the need for data sharing at national and international levels to support a debate on data quality assurance in survey findings.

Data sharing between government statisticians and the academic community can be mutually beneficial. With regard to official statistics, the public needs to see academic research based on national survey data as an extension and enhancement of official statistical information. This could increase public trust in official data and in the national agencies responsible for its compilation. In the process official data-collection, academic research and policy formation will benefit (Cook, 2003:7). The US government's policy to promote free access to government generated data is based on US policymakers' belief that such a policy is beneficial to the economic growth of the country. Empirical input from data collection is seen by policymakers as

necessary for the growth of the national economy (Weis, 2004:70). The free flow of government information is also seen as promoting good governance, as access to government information can be viewed as essential to promote “a more citizen-centred government” (Johnson, 2005:1).

Failure to support secondary analysis of survey data could mean this resource will be under-exploited, and opportunities for national growth, and for improving the quality of the survey research process through re-examination of data, will be lost (Musgrave, 2003:4). It makes scientific sense to utilise this data optimally, and organised structures for data preservation and data sharing can play an important role in this regard.

### **3.2.1 The Value of Data Sharing for Academic Research**

Ensuring the results of survey research are available for peer scrutiny promotes sound academic practice. Data sharing informs future survey work and reduces duplication, thus saving time and lowering the costs of research (Fienberg, Martin & Straf, 1985:5). The informal sharing of research data usually takes place among researchers in the same disciplines. Data sharing facilities can provide research resources across disciplines and countries and can therefore promote interdisciplinary and international survey research collaborations (Clubb et al., 1985:64). These partnerships expose researchers to new perspectives and methods, leading to further knowledge, to the benefit of research in general (Hedrick, 1985:130).

Sharing data reinforces transparent research, which is important to ensure its scientific validity. If the original research findings are made available for re-analysis or secondary analysis the validity of the original survey findings can be confirmed or denied. Primary researchers are restricted by the time-frame in which they work, as well as their nationality and the prevailing statistical techniques of the time, which could bias their interpretation of the data they collect. Making this data available on a long-term basis to secondary analysts could allow further interpretation of the data, to the advantage of academic research (Tanenbaum & Taylor, 1991:229-231).

Ready access to data for peers may also discourage dishonest research such as the deliberate forging of research findings (Hedrick, 1985:130). Errors in the original research can also be identified. These include errors arising from lack of rigour in the initial survey methods, as well as inaccurate conclusions based on low quality survey data. This re-examination could ideally

lead to improvements in data collection techniques and methods of analysis (Fienberg, Martin & Straf, 1985:10-12). Ultimately data quality could be enhanced, to the benefit of academic research generally.

### **3.2.2 The Educational Value of Data Sharing**

Existing datasets from completed surveys can provide valuable instructional material for university courses. These provide academics with a low-cost means of introducing their graduate students to the practicalities of survey design and analysis. It is possible that these students may in the future make useful contributions to survey methodology, based on early introductions to the field (Hedrick, 1985:131-2).

### **3.2.3 The Value of Data Sharing for Economic Growth**

The utilisation of research data has been advocated for sound economic policy-making, and to promote innovation to advance national economies. This is seen as a component of national information management usage to support development. Countries that do not create strategies for dealing with information are expected to be left behind in the global knowledge economy (National Research Council, 1995:19). Economic value can be derived from the free sharing of official statistics. Data that has been collected over an extended period of time, and has already been funded by taxpayers, can be utilised to the benefit of national economies if placed in the public domain.

Economic development is linked to a country's ability to promote applied research, and governments' commitment to invest in enabling infrastructures to facilitate knowledge utilisation for economic growth (Bugliarello, 1995:61-62). This belief in the value of research for the economic growth has translated in some countries into support for free access to publicly funded data sources. For example the United States' government promotes free and open access to taxpayer-funded information. This is in contrast to data archives linked to educational institutions in the US, many of which charge for data access. However, sharing of government data is not altruistic but rather reflects the belief of government policy-makers that the free sharing of official data will convey economic and social advantages to the countries concerned.

There is evidence to endorse the value of this type of data sharing for economic development. In Europe, where the sharing of public information has been formalised, government agencies

have recently begun to try and recoup costs by charging for official information. This may be short-sighted because the cost they wish to recoup is more than outweighed by the return on investment in data availability. The US spends approximately twice as much taxpayers' money on collecting and disseminating government funded public information as countries of the EU. To support their policy, the return on their investment in the provision of public data - in terms of economic growth and tax revenue - has been shown to be approximately five times larger than in the EU (Weis, 2004: 70-72).

A 2006 study on the commercial use of public information by the UK Office of Fair Trade found that more efficient access to public data for business could benefit the UK economy by about £1 billion a year. The research showed that government was often the only source of certain types of information – such as key administrative and geographical data – and if this data was made freely available to be utilised by commercial entities the resultant value-added products and services could have great benefit for the UK economy (United Kingdom, Office of Fair Trading, 2006:1-6). While this report refers to administrative data, exploiting the commercial value of national survey data could similarly promote innovation and economic growth.

### **3.2.4 The Value of Data Sharing for Good Governance**

Empirical data from survey research can be used for public policy formulation, with direct impact on national social and economic development. Public policy formation based on empirical data could lead to more informed decision-making by government authorities (Clubb et al., 1985:4, 10). Empirical social research can provide the raw material for evidence-based policymaking. The collection of national statistics via survey research to inform government policy can be seen as a vital component of governance. Where research feeds into public policy, re-examination of initial data is vital. Re-analysis of quantitative data can prevent inappropriate policy decisions based on inaccurate research, thus ensuring that policy formulation arises from sound survey research (Fienberg, Martin & Straf, 1985:4, 10).

Reuse of survey findings for different projects can be a cost-effective method of information-gathering for national planning. For this reason data sharing to further research to support evidence-based policymaking is an established practice in some European and North American countries. In these countries the sharing of empirical data is formalised by government policies and facilitated by National Statistics Offices (NSOs). These are assisted by SDAs, which are either components of national research support structures or units attached to academic

## SURVEY DATA ARCHIVES AND DATA SHARING

institutions. Examples of the former include the National SDAs in Europe. An example of the latter is the archive of the Inter-University Consortium for Political and Social Research (ICPSR) at the University of Michigan in the United States (Inter-University Consortium for Political and Social Research, [2009?]a). It is this belief in statistics as a valuable development resource that leads international donor organisations to fund capacity building projects and donate to the establishment of data management institutions in Africa (Mudesir Seid, 2006; Welch, 2007). Donor organisations place value on the collection and dissemination of sound statistical information to advance African societies. The role of data sharing in government policy-making in Africa will be examined in a later chapter on knowledge utilisation in African countries.

### 3.3 Data Archives and Data Sharing

Historically, data sharing among researchers has taken place in an ad hoc manner, and much valuable quantitative research has been unavailable to policy-makers and the wider research community. Despite the advantages of data sharing espoused by the academic community, many researchers are still reluctant to share their data. This is due partly to the technical and logistical difficulties of data sharing and to some extent to the dissuading motivations inherent in academic research, which is an environment in which exclusive access to original data can give researchers advantages over rivals in an academic field.

It is argued here that the establishment of SDAs to preserve and share data as a national resource provides a solution for many of the problems that occur when data sharing is only conducted on an informal researcher-to-researcher basis. SDAs can promote data sharing by removing technical and legal restrictions, and providing encouragement to and institutional pressure on researchers to share their data. SDAs facilitate data sharing through taking responsibility for the preservation and dissemination of data from social surveys after the original survey project is completed. These archives can also form part of regional infrastructures for the formal sharing of data. An example of this is the network of Survey Data Archives that exist in Europe, which, through their MADIERA portal provide researchers with multi-lingual access to any survey dataset available from any European country (Multilingual Access to Data Infrastructures of the European Research Area [MADIERA], 2009). An African example is the South African Data Archive (SADA) which provides access to its data holding for bona fide researchers in the region and internationally (South African Data Archive, 2009).

## SURVEY DATA ARCHIVES AND DATA SHARING

They can market data resources and provide facilities for formal training in data analysis and data preservation. SDAs can also play a quality assurance role with regard to the data that they make available for sharing.

With regard to the promotion of survey data as a teaching tool, SDAs facilitate data access for students, who may struggle to obtain important survey findings through informal channels. By undertaking to protect sensitive data, ensuring that available data is in user-friendly formats, and providing training in data analysis, SDAs allow even undergraduates to use quantitative data from surveys. For example, DataFirst at the University of Cape Town has signed confidentiality agreements with data depositors with regard to restricted data, and in this way ensures researchers at the university will have access to this type of data.

Data archives can act as intermediaries between data depositors and secondary analysts, and this provides for better research (Fienberg, Martin & Straf, 1985:13). By making the existence of datasets known to the wider academic community, SDAs can allow the creation of new datasets for secondary analysis by combining existing data files (Hedrick, 1985:129). Archive staff can facilitate ongoing interaction with the original data depositors which could ensure that existing datasets are combined and compared appropriately, to produce a relevant final data product.

Some early evidence exists for the advantages of the presence of skilled data archive staff to support data re-use. Lubanski in 1996 attempted to provide some empirical evidence for the continued funding of data support services at research institutions. In his research he quotes article content analysis studies indicating extensive growth in empirically-based journal articles to back his view that data repurposing will continue to escalate, and demand for data support services will increase (Lubanski, 1996:17). He details the results of a 1995 Fulbright Study undertaken to assess data support services at research institutions. The investigation involved interviews with researchers, data service staff and IT staff at seventeen research universities and eight research centres in North America, who were asked to assess the extent to which they utilise and value data services at their institutions. Opinions from data support staff were that there had been an exponential increase in demand for both survey data and data services. This research also showed that improved data support services correlated with increased research output, and researchers participating in the study expressed the view that data support was an essential component of a productive research environment (Lubanski, 1996:17-18).

## SURVEY DATA ARCHIVES AND DATA SHARING

This chapter examined the importance of empirical data for scientific research and teaching and as input for sound national decision-making for informed governance and economic growth. SDAs were investigated as institutions able to optimally facilitate the re-use of this data. In the next chapter another key role of SDAs is elaborated on. This concerns their capacity to support the quality dimensions of data products.

University of Cape Town



Photograph 3: The Norwegian Social Science Data Services (NSSDS) (Photo: NSSDS website, 2008: [http://www.nsd.uib.no/common/img/ill\\_nsdhus\\_large.jpg](http://www.nsd.uib.no/common/img/ill_nsdhus_large.jpg)).

[www.sociologia.sav.sk](http://www.sociologia.sav.sk)

Sociologický ústav SAV  
Klemensova 19  
813 64 Bratislava



Photograph 4: The Data Archive of the Institute of Sociology, the Slovak Republic (SASD). (Photo: SASD website, 2008: <http://www.sociologia.sav.sk>).

## CHAPTER 4: THE ROLE OF SDAS IN SAFEGUARDING DATA QUALITY

The previous chapter examined the primary role of Survey Data Archives which is to share official survey data for research purposes and to feed into effective national policymaking. This chapter looks at the valuable contribution these archives can make in assisting improvements in the quality of the data from national data producers and other survey organisations.

### 4.1 Definition of Data Quality and Quality Dimensions

The management of data quality in surveys is vital to satisfy the needs of those who will be using the final datasets. Data users include the data producer's original target group, as well as those who wish to use the data for repurposing. For this reason survey organisations should ensure in the first place that data quality targets (with regard to relevance, accuracy and timeliness, for example) are part of the stated objectives of survey organisations (US Census Bureau, Methodology and Standards Council, 2006a:1-2; Statistics Canada, 2003:12; Statistics South Africa, 2008:4-5).

Data quality can be defined as "fitness for use" (US Census Bureau, Methodology and Standards Council, 2006a:1). This definition usually uses key attributes to judge the quality of survey data. These include dimensions related to relevance, or the data's ability to meet users' information needs. Data users may be government agencies, academics, or private commercial clients. Accuracy is the second key dimension of data quality, here defined as "the difference between an estimate of a parameter and its true value" (US Census Bureau, Methodology and Standards Council, 2006a:1). Means of assessing this should be put in place at the design and implementation stage of the survey research project (Statistics Canada, 2003:87-88; Statistics South Africa, 2008:4).

Thirdly, data accessibility is important, that is, how easy the data is to find and use (US Census Bureau, Methodology and Standards Council, 2006a:1-2). The fourth dimension dealt with here is data timeliness, as indicated by the amount of time between the reference period of the data and its availability to the public. This always involves a trade-off between accuracy and cost (Statistics Canada, 2003:88).

The fifth dimension of data quality is *comparability*, judged by whether data from the same statistical operation can be compared at different times, between different geographical locations (international comparability) and between different “domains”, for example between industries, or household types (Rosen & Elvers, 1997:627; Statistics Canada, 2003:90). Comparability allows for the merging of disparate datasets (Cooper, 2005:2) and is dependent on the use of clear and appropriate classification systems, standard frameworks and stable survey methodologies (Rosen & Elvers, 1997:626; Statistics Canada, 2003:15;). The final dimension of data quality is its *interpretability*, which is dependent on the availability of useful documentation to support sound data analysis (US Census Bureau, Methodology and Standards Council, 2006a:2-4; Statistics South Africa, 2008:16; Statistics Canada, 2003:6-7).

### **4.2 Legislation and Guidelines Relating to Data Quality**

The extent of the benefits of data sharing will depend on the quality of the data shared. Country-specific legislation governing the collection and sharing of official statistics is mainly concerned with controlling misuse of the data relating, for example, to the protection of personal information (Niva, Sundgrun & Lyberg, 2003:24). However, laws such as the US Data Quality Act of 2002 do deal with other aspects of data quality. This Act requires federal agencies to issue quality guidelines ensuring the utility and integrity of the information they make available to the public (Center for Regulatory Effectiveness, [2002?]). Internationally the production and management of government statistics is guided by the United Nations Statistical Commission’s Fundamental Principles of Official Statistics (United Nations Statistical Commission, 1994). These principles assist official data producers in balancing the often competing objectives of relevance, accuracy, timeliness, cost and reporting burden, to achieve a quality survey product (Statistics Canada, 2003:4).

There is general awareness among survey organisations and principal investigators of the need for strategies to ensure the quality of their data products. Most NSOs undertake to use established data quality measures on their data products (Statistics South Africa, 2008:2). They undertake to conduct regular data quality evaluations and involve their data users in these assessments (Statistics Canada, 2003:58). Their policy documents emphasize the need to inform data users of all aspects of their research that could affect the quality of their data. This includes concepts and definitions used, as well as their choice of survey methodology and data protection techniques. Public access to this information is vital to enable data analysts to assess the usability of the data in an informed manner. Some NSOs provide official communication

channels for questions around the quality of particular datasets and undertake to obtain user feedback, for example in the form of user satisfaction surveys (US Census Bureau, Methodology and Standards Council, 2006c:1-2; United Kingdom, Office of National Statistics, 2009).

### **4.3 The Role of SDAs in Promoting Data Quality**

Where SDA staff form part of the principal investigation team, as is sometimes the case in research undertaken by universities, input can be provided to ensure that quality considerations form part of the project's initial data management plan. However, much quantitative social survey research is undertaken by NSOs in each country, where SDAs can provide no support for data management. However, SDAs can influence the final data product emanating from these government bodies firstly through their existence as public spaces for official and private data producers to deposit their data, and by providing incentives for them to do so. Secondly, the use of data archiving best practice by these archives can promote sound survey research and analysis methods on the part of data depositors and thus ensure the quality of the final data product. In this way SDAs can address impediments to the production of quality datasets.

Optimal methods of survey data archiving are covered in the literature from Europe and the United States. Best practice manuals emphasise the need to include archiving plans in the initial design of a survey (Humphrey, et al., 2000:2; Munoz, 2005:332). Funding agencies have acknowledged the value of this, and have begun to request that data archiving and data sharing plans be included in proposals for research funding (Inter-university Consortium for Political and Social Research [ICPSR], 2009:3). SDAs can provide expertise at an early stage to ensure the production of quality datasets. Once a dataset has been archived, these data archives can ensure the quality of the data is maintained and enhanced by supporting the various dimensions of data quality. These quality dimensions will be discussed in the following section.

#### **4.3.1 Ensuring Data Relevance**

Factors relating to relevance include the data's potential to address current issues, for example to provide empirical resources to improve public policy. Relevance may also refer to the utility of the data for the purpose of replicating and confirming earlier findings (Carley & Card, 2000:10). The relevance of statistical output to the user community may be difficult to measure, as it involves identifying the users of the data and communicating with them to ensure their data

needs are met, which can be a costly and time-consuming exercise (Lynn, 2004:576). However, this is necessary to ensure data producers understand the needs of their users, and users can understand how the data can be applied (Lynn, 2004:577). SDAs can act as intermediaries between data providers and data users to ensure the production of relevant and usable datasets, and their appropriate application. They are in a unique position to liaise with data users and data producers and identify the changing needs of secondary analysts, and how these can be addressed. Data Archive staff can provide feedback to data depositors on the usage and pertinence of their data. This could lead to improvements in the focus of national surveys, thereby enhancing the relevance of the data product. This has been the case in South Africa, where data archive staff are regarded by the National Statistics Offices as key stakeholders and invited to provide input for future government surveys.

### **4.3.2 Ensuring Data Accessibility**

There are several reasons why survey data may not be readily available to secondary users. Barriers to accessibility include the motivations of academia: principal investigators in survey research wish to have exclusive access to data to maximise their academic advantage. They have little interest in devoting resources to preparing datasets for the benefit of academic competitors. Research is motivated by rewards for publishing original findings, and investigators may withhold data for fear of pre-emptive publication by other academics, or inappropriate usage of their research findings (Fienberg, Martin & Straf, 1985:17). Technical and logistical hurdles preventing data accessibility include difficulties in identification and location of suitable data for secondary analysis, as information about survey data may not be readily available to researchers. Hardware and software incompatibilities may limit re-use of data that is available. Legal and ethical restrictions on data access relate to the need to ensure data confidentiality. The data archive network in Europe supports data accessibility by providing access via the MADIERA portal to the holdings of all European data archives (Multilingual Access to Data Infrastructures of the European Research Area [MADIERA], 2009). In Africa SADA and DataFirst support data discovery and ensure data access for researchers via their websites (DataFirst, 2009; South African Data Archive, 2009).

#### **4.3.2.1 Countering the Dissuading Influence of the Academic Reward System**

Because the primary task of survey researchers is data collection and analysis, and not data sharing, they may find the amount of time and money they need to spend to prepare their data

## SURVEY DATA ARCHIVES AND DATA QUALITY

for sharing unmanageable. Research funders have historically not made provision in their grants for the costs of data cleaning and metadata creation that will facilitate re-use of data from the research they sponsor (Fienberg, Martin & Straf, 1985:17-18; Rowe, 1976:99). Thus principal investigators may need to use their own time and financial resources to administer the data for re-use. If they have to provide constant feedback to users of their data, this may impinge on their research time. The administrative demands of preparing the data and documentation, advertising its existence, and providing information to secondary analysts may be too burdensome, and may prevent primary investigators from placing their data in the public domain (Clubb et al., 1985:63).

In the 1970s the Association of Public Data Users (APDU) sought to persuade survey researchers to release their data to the public. This was largely unsuccessful because academics focused on the advantages of exclusive “ownership” of their data. Researchers were also not prepared to invest resources in making their data available to fellow academics or the general public, and there were no sanctions from funding agencies or professional organisations for not doing so (Rowe, 1976:99). The advent of SDAs as organisations to facilitate data sharing saw new arguments with regard to control over survey data. Instead of providing a solution to problems of data ownership among academics, some SDAs sought to have control of large amounts of data through centralizing data collections. In Chapter 2, early US proposals for a central world archive were discussed: one of the reasons for resistance to the idea was that this was seen as an attempt to take ownership of data. These proposals were not supported because potential depositors of data feared loss of control over their data. (Scheuch, 1990:96; Scheuch, 2003:387).

Ownership and control of data continues to be an emotive issue. Some US organisations have used their superior funding and technical resources to harvest census and survey data from African countries, but have not placed their data in the public domain. An example of such an organisation is the African Census Analysis Project (ACAP), a project based at the University of Pennsylvania’s Population Studies Center. The project investigators do publish with invited co-authors from African institutions: however, not all the preserved data is available to researchers outside the Project (African Census Analysis Project [ACAP], 2009). It is assumed that the Project’s focus is on ensuring the continued existence of census material that may otherwise be lost. This can be seen as a worthy cause, but data archiving projects using donor funding could provide greater benefit to research by making their complete data holdings available to the wider academic community.

SDAs are dedicated facilities for data sharing, and can take the responsibility for data preparation away from primary investigators. Thus they can ensure access to data is not curtailed by the limited resources and different focus of initial investigators. They can also counter the dissuading factors inherent in the academic reward system where exclusive use of data allows researchers to produce papers derived from analysis of the data without having to share the findings which would dilute the impact of their research and therefore their academic rating. SDAs have assisted in bringing the issue of data sharing to the attention of funding agencies, and funders have begun to support data archives because they are seen as valuable resources for academic advancement, and for their role in promoting scarce quantitative skills, particularly in developing countries. An example of this is the funding allocated in 2000 for the establishment of DataFirst, a SDA at the University of Cape Town, by the US-based Mellon Foundation, in part to advance quantitative skills among graduate students in South Africa (Welch, 2007).

As already noted, some funding agencies have begun to require grantees to prepare their data for secondary use, and release it in a timely manner, as part of their stipulations for access to research grants (Inter-University Consortium for Political and Social Research, 2005:1). For example, in the UK ESRC award holders are required to deposit their data with the UK Data Archive (Schurer, 2008). This is a positive trend in encouraging data sharing among survey researchers, and is made possible by the existence of SDAs in many countries that fulfil the role of national survey data curating institutions.

### **4.3.2.2 Facilitating Data Discovery**

Researchers face the problem of locating information on the availability of datasets relevant to their research needs, and of accessing this data once they become aware of its existence (Hedrick, 1985:138). SDAs could assist data discovery through marketing their holdings and facilitating access to the datasets they house. International cooperation in the creation of online catalogues of data holdings by these archives has led to the creation of web-based national and regional catalogues for data discovery. For example, the Multilingual Access to Data Infrastructures of the European Research Area (MADIERA) Project provides online access to data in the Norwegian, Danish, Finnish, Greek, Swiss, German and UK SDAs (Multilingual Access to Data Infrastructures of the European Research Area [MADIERA], 2009). In South Africa both

DataFirst and the South African Data Archive (SADA) post catalogues of their data holdings on their websites (DataFirst, 2009; South African Data Archive, 2009).

#### **4.3.2.3 Supporting Technical Compatibility**

Data sharing is hampered in some countries by a lack of standards for data exchange. However, international technological standards are being developed to solve problems of technical incompatibilities. These include standards related to telecommunications, computer languages, storage media, terminology, and database management systems (Committee on Issues in the Transborder Flow of Scientific Data and National Research Council, 1997:29). These standards are often commercially-driven but can benefit academic data sharing. Modern SDAs are concerned with the promotion of these standards, which have made national and international data sharing possible. They also provide technical conversion capabilities for data sharing, which are lacking in informal data sharing networks (Clubb et al., 1985:64).

Problems related to the funding and control of information transfer via the internet can be barriers to data sharing. These arise because communication technologies developed for scientific data sharing, such as the internet and the World Wide Web, are no longer under the control of academic communities. The structure of the global information system and the legal framework controlling conflicts around access to data are now being defined by commercial and entertainment interests, which do not regard data-sharing as a priority (Committee on Issues in the Transborder Flow of Scientific Data and National Research Council, 1997:3). SDAs can exert an influence to ensure that academic interests are accorded importance in the global information sphere.

#### **4.3.2.4 The Data Quality/Data Protection Compromise**

SDAs can assist in ensuring that privacy and confidentiality issues do not restrict access to data. Privacy refers to control over access to an individual, while confidentiality refers to access to data about an individual, and agreements controlling this access (Lewis-Beck, Bryman & Liao, 2004:860). Protecting the confidentiality of respondent data is the main concern in data sharing, as privacy issues relate mainly to data collection, rather than the sharing of data once it has been collected. Ideally, confidentiality of data should be integral to any data management

## SURVEY DATA ARCHIVES AND DATA QUALITY

arrangement, and plans to protect confidentiality should be incorporated into the research design of a survey project (Lewis-Beck, Bryman & Liao, 2004:862).

Governments have established NSOs which are legally mandated to collect individual and household data to inform the creation and monitoring of government policies, through regular censuses and sample surveys. Information from these is generally made publicly available in reports or as aggregated data. To protect the confidentiality of the data collected, legally NSOs can only distribute data that does not identify individual survey respondents. EU legislation in this regard specifies that statistical data is confidential when it allows for the identification of statistical units which can lead to disclosure of information on individuals (European Union, 1997). Protection of confidential data is one of the fundamental principles espoused by government statistical agencies worldwide, as a necessary precondition to ensuring the trust of respondents, and therefore the accuracy and reliability of the data collected (Cook, 2003:1).

Recently, increasing demands have been made by researchers and policymakers for access to unit records from official surveys. Reasons for this are that these records offer great value for the investigation of social issues and the provision of solutions to social problems. For example, an important recent finding with regard to statistical data is that a truly accurate picture of the national economy of a country cannot be obtained from analysis of only aggregated data (Lane, 2003: 12). Thus access to unit records becomes important for research impacting on government economic and social policy. Often survey data linking responses to individuals is required to enable re-use of the data, for example with epidemiological studies, or panel studies (Fienberg, Martin & Straf, 1985:20) and restrictions on the use of these linkages can hamper research. Efforts to protect confidentiality can lead to other restrictions on the repurposing of official data. In the US, for example, the Freedom of Information Act of 1966 governs the usage of any data that does not remain under the control of the initial investigator (Fienberg, Martin & Straf, 1985:19-20). This type of legislation can curtail the useful sharing of research data if it is used inappropriately.

Primary investigators fear sharing of their data may lead to the breaking of confidentiality agreements made with survey respondents when they relinquish control over the original data. While concerns over confidentiality need to be taken into account with regard to data sharing, they do not provide a valid excuse for not sharing data. Researchers need to have concern for the best interests of both research participants and future users of their data (Taylor,

## SURVEY DATA ARCHIVES AND DATA QUALITY

1994:526). NSOs and other data producers need to find creative ways to provide open access to survey microdata for further research and to inform policymaking, while protecting the confidentiality of information entrusted to them (Lane, 2003:11). Several methods for the protection of confidential data have been developed, and SDAs can provide support for the implementation of these methods which allow data reuse. These are discussed in the following section.

Methods used by NSOs and other survey producers for protecting the confidentiality of data fall into two categories. The first method involves changes to the data at the data-creation stage, while the second restricts access at the data dissemination stage. The first method of protecting confidentiality involves making the data freely available, but altering the data to prevent identification of individuals in the survey. Various data masking techniques have been developed to prevent identification of individuals while preserving the integrity of datasets. Statistical methods such as the use of rounding techniques can be used to prepare data for distribution. Procedural methods include the elimination of identifiers from records (Boruch, 1985:99-100; Schouten & Cigrang, 2003:381-2). The advantage of anonymising data is that this enables the wider distribution of survey datasets.

The second strategy for the protection of confidential data is restricting data access. Two methods are used: on-site access, or remote access. Facilitating on-site access involves the provision of closed venues where researchers can access restricted data files. (Mackie & Bradburn, 2000:2-3). This ensures that data made available for research is only used for research purposes (Boruch, 1985:99-100; Schouten & Cigrang, 2003:382). Researchers accessing data at this type of facility sign contracts against misuse of the data. Examples include the data centre provided by the University of Michigan for researchers to access the United States Census Bureau's microdata, and Statistics Canada's research data centres, where sensitive datasets are made available under strict conditions of access.<sup>2</sup>

Remote access to survey microdata is provided by a few NSOs and research organisations. In these systems, data is provided via an internet connection, to enable researchers to undertake analyses on the data from their own workstations. In this type of data access, researchers submit queries using the query language of their chosen statistical analysis software. The NSO

---

<sup>2</sup> This information was obtained from a visit to the University of Michigan's restricted data facility in Ann Arbor and a Statistics Canada data enclave at McGill University, Montreal, in 2003 and 2007 respectively.

## SURVEY DATA ARCHIVES AND DATA QUALITY

executes the query through an automated system, and supplies output to the researcher, possibly in an anonymised form. The system includes various checks for breaches of confidentiality which could lead to misuse of the data (Schouten & Cigrang, 2003:382). Researchers utilising remote access systems usually need to sign contracts or obtain data use licenses authorising them to use the data at their work sites, with restrictions to prevent breaches of confidentiality, and the imposition of penalties for non-compliance with these requirements (Mackie & Bradburn, 2000: 34-35).

The advantage of remote access systems is that the security of the data is protected, as it remains on-site at the NSO or research organisation. At the same time this form of access obviates the necessity of researchers having to travel to research institutions to conduct research on the data (Schouten & Cigrang, 2003:382). NSOs providing remote access to their data include Statistics Sweden, whose remote access system provides internet access to data for researchers in any EU country (Söderberg, 2005:5). Secure data centres and remote access systems regulate access to the original microdata from censuses and sample surveys, and many researchers endorse these methods over modifying the data as a means of protecting data confidentiality (Mackie & Bradburn, 2000: 35).

As formal structures for data sharing, SDAs are able to support these data protection methods. They can accomplish this by assisting in the anonymising of datasets, for example by removing identifying information before making data available for secondary analysis. They may be able to access funding for this process if principal investigators are unable to fund this type of data preparation. SDAs could assist in limiting data analysis problems introduced by data altering techniques. These include the possibility of unintended effects on data modelling introduced by such methods. Elaborate data anonymising techniques could also require more advanced skills from data re-users, further restricting re-analysis or secondary analysis of the data. Skilled staff at Data Archives can assist with technical knowledge and report any data anomalies to data producers allowing them to re-examine their data perturbation methods.

The utilisation of centralized “safe centres” for the analysis of confidential national data is beneficial for researchers as these facilitate access to confidential unit records from surveys. However, accessing data in this manner can be inconvenient and time-consuming, and statistical agencies in some countries may be reluctant to provide this additional service (Mackie & Bradburn, 2000:10). SDAs can facilitate wider usage of restricted access datasets by acting as

## SURVEY DATA ARCHIVES AND DATA QUALITY

secure facilities for this purpose, as is the case with archives in North America and Europe (Dunne & Austin, 1998:19-21; Inter-University Consortium for Political and Social Research, 2009:32).

These facilities can play an important role in mitigating the confidentiality concerns of respondents and principal investigators. Considerations around data confidentiality are compounded in cross-national research (Taylor, 1994:523). However, issues of international concern can be effectively handled by international networks of SDAs. These networked archives can manage and monitor adherence to ethical principles in complex cross-national research, while standardizing data and access procedures (Taylor, 1994:525).

Archives can be a point of liaison between official statistical agencies and the users of their data. Data Archive staff can work with data producers to investigate the impact of confidentiality measures on data quality (Lane, 2003:18). However, to accomplish this effectively they need to be familiar with legal and institutional restrictions on data repurposing and any changes to these over time. Data Archives need to produce policy documentation on confidentiality protection, so as to inform data users on the laws and policies governing this aspect of data usage. At present this is not the case with all SDAs, but this is a vital role that these organisations can play in facilitating data sharing (Mackie & Bradburn, 2000:40).

### **4.3.3 Promoting Accurate Survey Data**

SDAs can support the production and use of accurate survey data through encouraging greater rigour in survey research by principal investigators who will be aware that their data will be widely analysed if it is available in these archives. These institutions can also play a pro-active role in checking for anomalies in archived data.

#### **4.3.3.1 Supporting Sound Research Methods**

The existence of SDAs can prevent restrictions on data sharing that may be motivated by the need to disguise poor quality research. Lack of concern for academic rigour in research may explain the reluctance on the part of some researchers to share their work. Primary investigators sometimes resist sharing their data because they fear criticism of their work, which could include exposure of their inadequate research methods or rebuttal of findings (Fienberg, Martin & Straf, 1985:17). However, this is not a legitimate reason for withholding access to data, as academic rigour on the part of initial investigators should preclude invalid criticism of

their methods and findings. Critical commentary is part of academic debate and if principal investigators have confidence in the soundness of their research they should not fear close examination of their primary data. The existence of SDAs may encourage quality research as primary investigators will be aware that these repositories will facilitate wider access to their datasets and this will increase the possibility of careless work being discovered.

### **4.3.3.2 Error Reporting and Correction**

Informal data sharing arrangements do not necessarily allow for improvement in the quality of existing datasets. Primary investigators often do not take kindly to having errors in their survey methods and findings identified. Unless formal channels for data sharing exist, error reporting may not lead to corrections in datasets and improvements in survey methods (Hedrick, 1985:139). SDAs provide a formal structure for the identification of errors in data and documentation. This should lead to, at best, their correction or, at worst, the publishing of provisos on the limitations of a particular dataset. The data quality statements of NSOs encourage regular and public revisions of their findings (Statistics South Africa, 2008:21) and SDAs can assist in the public distribution of these findings. For example, the ICPSR has since the 1970s been providing feedback to its membership on errata in their datasets and their correction (Memo from ICPSR staff to Official Representatives, 1976:7).

Quality control of national datasets is vital. Census and survey data inform policy-making and development action and thus need to be accurate and comparable. Principal Investigators are responsible for post-collection quality control procedures, such as editing and data review and correction. However, liaison with data producers and data users is necessary to correct errors found in the final data product. SDAs can take on the role of “data detectives” to assist in identifying errors in survey datasets. Consultation with national data producers could enable SDA staff to identify the causes of inaccurate data, such as coding errors. Collaboration with analysts and principal investigators could lead to datasets being revised and corrected. Some quality control measures may lead to the discovery that certain datasets are not fit for use, or are only usable for certain purposes.

An example of “data detective” work being undertaken is the quality control role played by DataFirst in investigating the usability of datasets produced by the South African NSO, Statistics South Africa (StatsSA) as part of the archive’s Data Quality Project (DataFirst website, 2009). Errors found are brought to the attention of StatsSA who provide new data and metadata files

## SURVEY DATA ARCHIVES AND DATA QUALITY

which are made available on the DataFirst website. National data producers should advertise errors and their correction as widely as possible to prevent inappropriate usage of survey datasets and the consequent publishing of dubious research findings. However, in countries where national data producers do not have procedures in place to inform the public of data errors, SDAs can take responsibility for investigating and reporting any data anomalies found at the post-dissemination stage.

Elucidating information can be vital for enabling the secondary analysis of some datasets. Often the initial sampling for a survey and the data-collection and data capturing are carried out by different organisations. Bearing in mind the number of agencies that may have input into the data product before the final dataset is received by the archive, the importance of being able unambiguously to trace the provenance of a dataset cannot be overemphasised. Sampling, enumeration and data conversion errors at this stage are beyond the control of the SDA that finally receives and houses the survey findings. However, the role of the SDA is to ensure provision of documentation on problems with datasets resulting from errors in sampling and data collection or conversion at the pre-acquisition stage and to control for errors that occur during the acquisition and archiving stages (Economic and Social Data Service [ESDS], Rural Economy and Land Use Programme Data Support Service, 2006:6).

### **4.3.4 Promoting Timely Data Release**

Survey data needs to be as current as possible to be useful, and for this reason data producers strive to meet reasonable release deadlines. NSOs may delay the release of datasets to undertake extra data cleaning activities. However, a compromise must always be reached between the timeliness of data releases and data quality, as reducing data collection and data cleaning periods can adversely affect the accuracy of the survey data (Rosen & Elvers, 1997: 626). Other obstacles to timely data release relate to the fears of loss of control over data expressed by the initial researchers and their need to extend embargo periods on data to maximise academic rewards for their research efforts (Fienberg, Martin & Straf, 1985:17).

Different types of survey research projects may also require different time-scales with regard to the release of final datasets for re-use. For example, investigators in longitudinal studies often delay data release until several waves of the study have been completed. Justifications for this relate to the increased time period needed to curate the vast amounts of data that are collected in such studies, and the added vigilance required by investigators to ensure the confidentiality of

## SURVEY DATA ARCHIVES AND DATA QUALITY

data in longitudinal research. The need to balance the property rights of data creators, the data access needs of secondary users of the data, and the confidentiality requirements of data subjects can result in delays in the public release of data (Durant and Menken, 2002:44).

However, some data producers may impose unreasonably lengthy embargos on the information they collect. This is a controversial issue in social research. Although most researchers would agree that open access to data from survey research can benefit the research community, the dissuasive influence of the academic reward system makes initial investigators reluctant to immediately share their work. The initial investigators in any social survey take the view that their primary role in the original research should mean they get first publication rights and they put systems in place to ensure at least temporary ownership of the data from their surveys (Dunne & Austin, 1998:1). However, foreign organisations conducting surveys in African countries may embargo the results for unjustifiably long periods to extend the publishing opportunities afforded by exclusive access to the data.

Viewed cynically, long embargo periods on data obtained in Africa may reflect the need for foreign academics to access new sources of survey data and publishing opportunities denied them in their own countries, where national datasets are more readily available to researchers and may have already been analysed extensively. The practice of embargoing survey results for years hampers access to data and can often be objected to on the grounds that publicly funded datasets belong in the public domain. Although initial researchers wish to maximise the advantages of conducting the original survey research, exclusive usage for long periods cannot be justified on these grounds. There are informal rewards within the academic community for producing and sharing quality data which can counterbalance losing publishing time with exclusive data. These include respect from academic peers for contributing to research and teaching, as well as promotions resulting from a sound reputation for the general advancement of scientific inquiry (Fienberg, Martin & Straf, 1985:29).

Issues of academic credibility are also relevant here. With embargoed data, there can be no independent assessment of the published work of initial investigators. Making the original data available to the research community supports sound academic research, through promoting the possibility of extensive peer review of this type of data. It is understandable that principal investigators are accorded first analysis of data collected from research initiated by them. There is a certain amount of consensus in the academic community regarding the right of initial investigators to exclusive access to microdata until publication of the initial report of their

findings. However, withholding access to data until all research on the findings has been exhausted should not be the norm with publicly funded survey research.

SDAs can minimize the embargo periods on data by assuaging the fears of initial investigators with regard to loss of control over their datasets. Archives can take responsibility for ensuring controlled usage of the data, for example for research purposes only. Archive staff can sign agreements with regard to data access, and take responsibility for ensuring researchers use the data appropriately.<sup>3</sup>

### 4.3.5 Supporting Data Coherence

This quality concept relates to sets of data, and refers to general coherence, that is the ability to combine the statistics with other data sources, for example administrative records. It also refers to how easily survey datasets can be compared with other surveys, over space (different geographical locations and domains) and over time (creating a statistical time-series) (Rosen & Elvers, 1997:626). Quality data ensures changes indicated are actual and not related to deviations in survey methodology or unreliable measurements. The creation of comparable data is only possible if data providers adhere to international standards with regard to definitions, concepts and methods (Rosen & Elvers, 1997:627).

In some countries official data producers lack the infrastructure to ensure that standards are adhered to in a consistent manner. This can lead to problems with data comparability and version control (Vahed, 2005:2). SDAs can work with NSOs to help maintain and enhance data quality by initiating and supporting the use of international standards for survey data. Survey data archiving best practice takes into account the nature of the final data product. Steps towards creating a quality data product include ensuring that questions, sampling procedures and classifications adhere to harmonisation principles, which relate to international standards of survey design to allow for comparability between surveys (Meltzer, 2003:3; Verma, 2002:3). Data Archives can assist with this process.

DataFirst, at the University of Cape Town in South Africa, plays a role in fostering the quality of official South African data through their Data Quality Project, which promotes an understanding of the strengths and limitations of this data. Working with Statistics South Africa, the unit

---

<sup>3</sup> V. Ramachandran, e-mail message to author, 24 January 2006.

provides information on anomalies and restrictions related to the data from the national data producer, to support the appropriate usage of this data (DataFirst, 2009).

### **4.3.6 Enhancing Data Interpretability**

Data interpretability is a function of good supporting documentation, that is, metadata (descriptions of data files and coding) and other explanatory documentation. Even if initial investigators do share their data, secondary analysts need to be able to understand it enough to reuse it. Supporting documentation for data may be inadequate because initial researchers have few resources to fully document their findings. Poorly documented research has limited value as a research tool for secondary analysis (Clubb et al., 1985:55-56). Inaccurate or inappropriate usage of data can be mitigated by the provision of detailed information with each dataset produced.

#### **4.3.6.1 Providing Value Added Input to Documentation**

Ideally, supporting documentation should include both explanatory and contextual information on the survey dataset described (UK Data Archive, 2002:2). The documentation provided with datasets should also include information on the strengths and limitations of the data (Statistics Canada, 2000:3). Data managers may need to create additional metadata to enable sound secondary analysis of the data. International standards for the creation of metadata, such as the Data Documentation Initiative (DDI), have been developed for this purpose, (Data Documentation Initiative, 2009).

Good metadata should be comprehensive enough to allow secondary users to understand the restrictions which may apply with regard to analysing the data (Musgrave, 2003:8). This will counteract the problem of primary researchers withholding complex datasets for fear of inappropriate use by analysts with limited statistical skills. SDA staff can apply their unique skills and the advanced technology available in their archives to create metadata files, digitize documents and generally ensure supporting documentation is kept current and accurate to optimize data analysis. Facilitating data sharing in itself should encourage improved data interpretation, as researchers working on official statistical data are often able to supply feedback to national data producers on weaknesses in data documentation, leading to improvements in the future documentation of official data.

## SURVEY DATA ARCHIVES AND DATA QUALITY

There are many pitfalls to creating comprehensive and useful documentation to support secondary analysis of survey data. One of these is the problem of outdated technology. Documents may be stored in formats that become obsolete. For example the October Household Survey (OHS) conducted by Statistics South Africa in 1995 was released with the questionnaire in the software program WordPerfect. Soon after the release of the survey this software programme became obsolete, and thus no readable digital version of the questionnaire existed. The use of this survey dataset was hampered by incomplete documentation resulting from incompatible technology. SDAs can have positive input here, by using the technology at their disposal to convert files to other formats before the current ones are unusable. In this case DataFirst was able to digitise an original hard copy version of the OHS questionnaire and provide this file to researchers via the DataFirst website.

University of Cape Town

#### 4.3.6.2 Preventing Archiving Errors

Dataset authenticity and integrity need to be ensured by the SDA, that is, systems need to be put in place to ensure data authenticity – the quality of genuineness and trustworthiness of the data, and data integrity - the state of being whole, uncorrupted and unchanged (National Library of Australia, 2003:147). SDAs can add value to datasets by working with the original data producers to create elucidating documentation on the structure of data files. They should be wary, though, of adding to the confusion. Inaccuracies may be introduced as a result of archiving errors once a dataset has been acquired by an archive. Problems arise here because digital materials can be altered easily and inadvertently, through the data management process. The processing of complex datasets to prepare them for archiving inevitably results in changes to the data and document files. These generally provide value-added material, but careless archiving can have disastrous results.

To give an example: SADA disseminates datasets from StatsSA and provided an early household survey, the October Household Survey of 1994, to both local and international researchers. The questionnaire for the 1994 October Household Survey was retyped by SADA staff for inclusion in the dataset. As this was a seldom-used dataset it was twelve years later that an observant economist at the University of Cape Town noticed the codes for race had been transposed in the retyped questionnaire, making nonsense of any race related data from this survey (Wittenberg, 2006:766-768). Data managers at DataFirst were able to liaise with SADA and Statistics South Africa regarding the problems with this dataset in order to make the necessary corrections.

This chapter dealt with ways in which SDAs assist in fostering the quality of survey datasets and concludes the first part of the study, which focused on aspects of international data curation. Part two deals with data curation in Africa, through an overview of historical developments and current practices, an investigation of obstacles to effective data management on the continent, and suggestions for improvements in this regard.

**PART TWO: SURVEY DATA CURATION IN AFRICA**

University of Cape Town



Photograph 5: The National Institute of Statistics and Economic Analysis (INSEA) in Benin (Photo: INSEA website, 2008: <http://www.insae-bj.org>).



Photograph 6: The Institute of National Statistics and Economic and Demographic Studies (INSEED) in Chad. (Photo: INSEAD website, 2009: <http://www.inseed-tchad.org>).

## CHAPTER 5: A SURVEY DATA CURATION HISTORY OF AFRICA

This chapter covers developments in the management and sharing of survey microdata on the African continent. Appendix C provides a timeline of major events in this sphere, some of which are further elaborated on in the paragraphs below. The African data curation index provided as Appendix E of this study can be consulted for an overview of the data curation attributes of African NSOs.

In Western Europe official information policies have been formulated to encourage the efficient collection of information, including statistical data, by government agencies. Information is viewed by policymakers as a national resource, and its optimal utilisation is encouraged for effective governance and economic growth, leading to a situation where:

Information is one of the few non-scarce raw materials for the economy of Western Europe. The timely and inexpensive availability of information makes it a motor for innovation and therefore a driving force for economic and social development (van Rosendaal, 1984:15).

In Africa data collection through censuses and sample surveys and the management of this data for reuse is constrained by several obstacles. There has been a substantial amount of African data collected over the years, mainly by African NSOs (Dupriez, 2008a:2). However, these organisations have limited financial and staff resources to curate this microdata and ensure its long-term availability. Consequently many African data producers do not follow international best practice with regard to survey data management or share the microdata from the surveys they conduct (Regional Reference Strategic Framework for Statistical Capacity Building in Africa, 2006:131).

However, with growing emphasis on the importance of statistical data as a national resource for scientific investigation to support innovation and sound national decision-making, some African leaders have begun to support the production and preservation of quality statistics for reuse as a planning tool. Government accountability in the region has become tied to policymaking based on evidence collected through survey research which is becoming an imperative to support political legitimacy. Effective government planning is increasingly seen to depend upon sound policy analysis by researchers utilising survey data, which enables them to confirm and extend the findings of government statisticians. The role of the research policy interface, in which sound research by academics enables effective government planning, is increasingly seen to depend

on access to original microdata files for researchers (Africa Symposium on Statistical Development 2006b:5; African Union, 2009:1).

The needs of international development organisations have also changed data management practices on the continent. These agencies require country-level data to monitor their development projects in the region. Development agencies, in partnership with African governments and regional organisations, have over the years provided funding to NSOs for the production and dissemination of national statistics, the use of such statistics to formulate national policies, and data repurposing for further research (Achikbache, et al. 2002:159; Ching'anda & Ntozi, 1998:1-2; Onsembe & Hie, 2004:1-3).

Recently, another data curation initiative has entered the African arena. This movement advocates the proper curation and free sharing of all research that is publicly funded, including survey research. The tenets of this initiative are articulated in the Organisation for Economic Cooperation and Development's (OECD) 2004 Declaration on Access to Research Data from Public Funding which advocates the sharing of data for international and regional co-operation in research. The initiative is linked to the establishment of data curation organisations as part of international and regional research support infrastructures and has led international development agencies to encourage African governments to create enabling mechanisms for the placing of government sponsored research data in the public domain (Organisation for Economic Co-operation and Development, 2007:1-13). ICT infrastructure development is also promoted in this regard as a necessary prerequisite for data management, including research data sharing.

### **5.1 The Colonial Period**

The findings of surveys in the colonies in Africa were published in statistical yearbooks and this practice continued until the independence of these African countries (Gervais & Marcoux, 1993:389-390). There was little emphasis on the preservation of the original records of these early African surveys, and no attempt was made to organise these for re-examination by future researchers. As a result, much valuable early African survey data has been lost as a research resource. For example, in 1989 a data restoration project initiated by the Programme Population et Developpement au Sahel (PPDS), at the University of Montreal, Canada, set out to collate and classify statistics from surveys conducted in Francophone African countries between 1960 and 1965. They found approximately 8000 pages of tables, most un-analysed and

lacking their original metadata, stored in the cellars of the Office de la Recherche Scientifique et Technique d'Outre Mer (ORSTROM), in Paris (Gervais & Marcoux, 1993:386).

### **5.1.1 International Support for African Data Curation**

From its establishment in the 1940s the World Bank established projects aimed at supporting statistical development in the region, for example, by providing technical assistance to African NSOs for the collection of national accounts data during this decade (Tulya-Muhika, 1990:556). In the 1950s the World Bank established field survey operations in African countries for the collection of socioeconomic. These surveys were mainly concerned with prices in the urban areas and with agricultural structure and production in the rural areas (Booker, Singh & Savane, 1980:2). International efforts, however, focused on support for *data collection* in Africa rather than the overall management of African survey data (Gervais & Marcoux, 1993:389).

Initiatives for data archiving were needed on the continent. The United Nations Economic Commission for Africa (UNECA) was established in 1958 and, as an organisation mandated to foster and coordinate the economic and social development of its fifty three member states in Africa, UNECA has been in a good position to nurture knowledge utilisation on the continent. The organisation's legislative body is the Conference of African Ministers Responsible for Economic and Social Development and Planning, established in 1974. This Ministerial Conference initiates the activities of UNECA and reports on these to the UN ECOSOC (United Nations Economic and Social Council (Economic Commission for Africa, 2008). UNECA functions as a key coordinating body for projects aimed at encouraging data production and management in the region. However, its coordinating role in this regard has not been supported by international development organisations or African governments consistently enough to facilitate regional advancement in the curation of African data resources (Lehohla, 2008:14-15).

### **5.1.2 African Data Curating Institutions in the Colonial Period**

Unlike in Europe and North America, data curation in African countries takes place almost exclusively within government structures. Most data collected and archived in Africa is official data. Any study of data curation in Africa will therefore need to focus mainly on the management and sharing of data within African National Statistical Systems (NSS). The structure and function of African NSS are influenced by their colonial past. Most NSOs in Africa were originally colonial records offices established to collect population and vital statistics data

as a by-product of colonial administrative functions and for revenue purposes (Tulya-Muhika, 1990:556).

Some social data on African populations was collected during the colonial period. For example population data was collected in Francophone West and Equatorial Africa during the 1950s by the French *Service Colonial de Statistiques* (SCS), utilising new survey sampling techniques introduced to France by American statisticians (Gervais & Marcoux, 1993:386-389). However, data collection for development planning was not part of the agenda of colonial administrations. NSS in Africa inherited a decentralized structure from colonial administrations, with NSOs as their key data curation agencies responsible for coordination of data collection and data management within these systems (Ching'anda & Ntozi, 1998:235; Eele, 1989:432-433).

At the close of the colonial period therefore, African statistical systems showed the characteristics that were to provide ongoing obstacles to effective data management on the continent. National statistical departments in African countries were underdeveloped and poorly managed, and data production was limited (Onsembe & Hie, 2004:1-2). These institutions also reflected the characteristics of colonial statistical systems in their failure to integrate data collection into government planning processes. NSS were not seen as essential components of governance, and therefore received little financial or logistical support from colonial administrations. Thus African NSS, and particularly NSOs within these systems, have struggled for relevance and legitimacy, and have historically found little encouragement from African governments for their role as curators of national data for economic development and policymaking.

### **5.2 The 1960s**

The 1960s saw an increase in the demand for empirical data in African countries as post-independence governments realised the need for national statistics to assist with development planning (Kiregyera, 2001:1). However, data collection in Africa through survey methods was plagued by problems during this period. Surveys were not regular or timely but rather conducted on an ad hoc basis, and data collection agencies in the region lacked uniform standards with regard to coverage and concepts. Consequently African governments were unable to use existing data to assess the situation in their countries, create suitable policies or

monitor policy outcomes (International Labour Office [ILO], Policy Integration Office, Bureau of Statistics, 2004:1).

### **5.2.1 Stakeholders in Coordination for Data Development in Africa**

To strengthen data management systems in Africa, the United Nations and other multilateral and bilateral agencies provided funding and logistical support for data preservation and dissemination in the region, in partnership with regional and national organisations. International and regional efforts to advance data curation in Africa adopted two key strategies. Firstly they focused on the development of country infrastructures to affect data curation in the region. This involved building institutional infrastructures for data management to advance the quality of African data to ensure its usability, and support for technological infrastructures for data preservation and data exchange. Secondly, donor organisations supported human resource development to provide countries with data curation skills. These two strategies, in various guises, remain at the heart of international programmes aimed at fostering African data management to the present day.

Later these programmes included support for *regulatory infrastructures* in African countries, and advocacy around data use for policymaking to promote sound governance. These latter initiatives took the form of assisting African governments in the formulation of enabling policies and appropriate legislation for data curation (Kiregyera, 2001:2-3; Regionalization of social sciences in Latin America, Asia and Africa, 1973:559). Initially ad-hoc, international sponsoring of sound information management in Africa has progressed to more coordinated programmes aimed at effective data curation for better governance and economic growth in countries of the region.

### **5.2.2 Improving Institutional Infrastructures to Advance Data Quality**

In the 1960s institutional capacity-building to advance the quality of the data produced by African countries came from multilateral agencies and foreign NSOs. This focused on African NSOs as key institutions within African NSS (Ching'anda, 1999:1). Statistical advancement strategies included the provision of training for and technical assistance to NSO staff and funding for equipment and census and survey programmes. These strategies mainly promoted data collection and data usage by African NSOs, through the United Nations Development Programme (UNDP). However, there were some attempts to build institutional capacity for data

preservation and data sharing on the continent at this time. These were part of general attempts to support African research institutions (Kiregyera, 2001:2-3; Regionalization of social sciences in Latin America, Asia and Africa,, 1973: 559-560).

### **5.2.3 Support for Technological Infrastructures in Africa**

During this period UNECA established its National Information Systems (NATIS) programme, mainly promoting libraries in African countries. Later, this project became part of UNESCO's General Information Programme (PGI), which supported information policies and information systems in Africa, and promoted regional co-operation in information management for development (Yumba, 2002:239).

### **5.2.4 Human Resource Development**

Training for data management in the 1960s was marginal and undertaken only in the context of the training of statisticians, usually for government service. Statistics personnel in post independence NSOs were trained in the UK or France, and were therefore not necessarily well equipped to deal with local conditions (Tulya-Muhika, 1990:557). To solve the staffing problems of African NSOs the 2nd Session of the Conference of African Statisticians initiated the African Statistical Training Programme (ASTP) in 1961 (Ching'anda & Ntozi, 1998:236). This programme provided funding for the creation of regional Statistical Training Centres (STCs) in African countries.

Three STCs were established to provide statistical training in Francophone African countries - ENEA (National Institute of Applied Economics) (Dakar, Senegal), ENSEA (National Higher Institute of Applied Statistics and Economics (Abidjan, Côte de l'voire) and ISSEA (Sub-Regional Institute of Applied Statistics and Economics (Yaounde, Cameroon). Two STCs provided training in Anglophone African countries: EASTC (East African Statistical Training Centre) (Dar-es-Salaam, Tanzania) and ISAE (Institute of Statistics and Applied Economics (at Makerere University, Kampala, Uganda) (Muba, 2004:2; N'Guessan & Chitou, 2006:152-5; Nyanzi, 2004:2).

ENSEA was founded in Rabat, Morocco to advance statistical skills in North Africa, along with the Cairo Demographic Center, established in Egypt in 1963 (MEDSTAT II, 2008:25). Contributions by STCs to practical data work were limited as they were geared to a theoretical

approach to statistics, and they were not geared to new data needs. The curricula of these training institutions placed very little emphasis on data management (N'Guessan & Chitou, 2006:155-156).

### **5.3 The 1970s and 1980s**

The 1970s and 1980s saw a decline in the data capacities of African countries, due largely to the low priority African governments placed on the collection and analysis of statistics to support economic and social policies (Ching'anda, 1999:1). There was a general lack of awareness by government planners of the value of empirical data to inform sound decision-making. Political instability also led to problems of legitimacy among African leaders, adding to their reluctance to share official statistics, for fear of revealing their political inadequacies. This crisis of legitimacy was reflected in state organs, including NSOs, as statistics were manipulated by some African governments to exert control over citizens and justify existing policies (Lehohla, 2008:6-8). NSOs also suffered budget cuts due to economic declines in African countries during this period. Funding shortages were coupled with staff shortages and the poor management of statistical institutions within African NSS, resulting in a general decline in statistical services on the continent.

#### **5.3.1 Stakeholder Coordination**

However, during this period a few information management projects were initiated by regional and international development agencies working in partnership. These promoted the effective use of information resources for national and regional growth in Africa and the building of African research capacities, as well as good practice in information management on the continent (Ching'anda, 1999:1-2). For example, the African Census Programme and the Statistical Development Programme for Africa provided resources for the Statistical Training Programme for Africa (STPA) to advance statistical capacities in the region.

International surveys, which included surveys of African countries, served to provide most African data during this period, and some established sound data management and data dissemination practices for surveys conducted in the region. An example is the World Bank's Living Standards Measurement Study (LSMS) programme. This was initiated in 1980 to provide "a better quantitative basis for the design and monitoring of development policy" (Chander, Grootaert & Pyatt, 1980:1). This purpose refers to the World Bank's own information needs, but also relates

to the data requirements of decision makers in African countries for government planning (Kiregyera, 2001:1-2).

### **5.3.2 Support for Institutional Infrastructures to Advance the Quality of African Data**

The quality of data produced by African NSS deteriorated during the 1970s and 1980s because the African economic crisis and the low priority African governments placed on statistics meant NSOs were chronically under-funded in this period (Kiregyera, 2001:1). The United Nations Fund for Population Activities (UNFPA) (later the United Nations Population Fund) played a key role in this regard for three decades, from 1970 to 1993, by supporting census programmes and promoting proper data management in the region, and assisting co-ordination between African governments and their regional and international development partners (Onsembe & Hie, 2004:1-3). The UNFPA established the African Census Programme in 1969 to provide funding and technical transfer to those African countries participating in the centennial census rounds of 1970, 1980, 1990 (United Nations Economic and Social Council, 2001:43). These three decades of censuses resulted in extensive data being produced by participating countries (Booker, Singh & Savane, 1980:2).

Some work on developing data quality was undertaken at this time. In 1973 the Conference of African Statisticians proposed the establishment of permanent field survey facilities in participating African countries. This included the provision of trained staff to be employed in African NSOs. From 1977 the ECOSOC of the United Nations Statistical Commission (UNSC) supported this programme, known as the African Household Survey Capability Programme (AHSCP). When the first working group of the AHSCP met in 1979 one of their concerns was the provision of comparable survey data on households in African countries. To this end they emphasised the use of internationally standardised concepts, definitions and classifications at NSOs to ensure the consistency of African surveys, and proposed a common set of core questions for African household surveys conducted by the World Bank (Booker, Singh & Savane, 1980: 8-11).

### **5.3.3 The Development of African Technological Infrastructures for Data Curation**

## A DATA CURATION HISTORY OF AFRICA

During the 1970s and 1980s UNECA advocated for the effective management of information for the achievement of Africa's development goals. Changing technologies, which ushered in the worldwide information economy, led this agency to initiate a programme promoting the use of information and communication technologies (ICT) to optimise knowledge utilisation in African countries. This focus originated from the belief that, although these technologies are no more than enabling tools in the development process, a solid ICT infrastructure is a necessary requirement for the effective management of information to inform development policies and affect economic growth (National Information and Communication (NICI) e-strategies, 2007:xiii).

In 1980 UNECA set up the Pan-African Development Information System (PADIS) as a co-operative regional information management project focusing on the development of information resources and information networking in the region. The PADIS project organised a series of seminars on the formulation of National Information Policies (NIPs) in African countries. These seminars were conducted from 1987 to 1991, with limited success. However, they served to initiate the international community's ICTs for Development (ICT4D) strategy to promote the use of modern information technologies for data curation to support growth in the African region (Yumba, 2002:239-240).

### **5.3.4 Human Resource Development for Data Curation - 1970s and 1980s**

During the 1970s the UNFPA founded demographic training and research centres in African countries, to promote statistical skills in the region. The Regional Institute for Population Studies (RIPS) was established at the University of Ghana in 1972 ((United Nations Population Information Network [POPIN], Regional Institute for Population Studies [RIPS], [2003?]). The Cairo Demographic Center was set up in Egypt and the Institute of Demographic Training and Research (IFORD) was established in 1971 in Cameroon at the University of Yaounde. However, these institutions did not undertake data management training except in the context of general statistical instruction (Lehohla, 2008:5).

UNECA continued to conduct workshops and meetings aimed at inculcating data quality standards among statistical practitioners in the region, with variable results. A 1977 UNECA assessment of the international community's ASTP found it had limited success in nurturing statistical skills in the region (Tulya-Muhika, 1990:557). In 1978 therefore, UNECA introduced the STPA as a new statistical training initiative (Ching'anda & Ntozi, 1998:236-237). This was

later extended and placed under the UNDP's Statistical Development Programme for Africa (SDPA) (Tulya-Muhika, 1990:557).

UNECA's statistical training programmes supported human resource development for statistics in African countries from 1978 to 1993, when funding from the UNDP ceased (Ching'anda, 1999:1; United Nations Economic and Social Council, 2001:43). This assistance involved financial and logistical support to the regional STCs (Tulya-Muhika, 1990:557-558). Statistical training conducted by the STCs included degree-level training at regional and national level, in-service training at national level (mainly in Anglophone African countries), post-graduate and specialized training and workshops and seminars. However specific training in data management was not provided as part of the programme.

### **5.4 Developments in African Data Curation in the 1990s**

By 1990, concerned about the weakened state of their statistical systems, the Economic Planning Ministers of UNECA member states mandated the African Development Bank (ADB) to undertake a detailed assessment of the NSS of African countries. On the strength of the ADB findings the Ministers developed a regional data support programme, the Addis Ababa Plan of Action for Statistical Development in Africa in the 1990s (AAPA) (Kiregyera, 2001:1-2).

#### **5.4.1 Stakeholder Coordination in the 1990s: The Addis Ababa Plan of Action for Statistical Development (AAPA)**

The AAPA gained impetus from political developments on the continent during this period. Moves toward democracy in many African countries at this stage, with a concomitant focus on government accountability, led to greater demand for empirical data from African governments for planning purposes. Statistics became a means of holding governments to account and the conducting of a census came to be seen as evidence of attempts at democratic rule in African countries (Lehohla, 2008:4). For example the 2005 Comprehensive Peace Agreement (CPA) ending the conflict in Sudan included the requirement for a Census to be undertaken in that country in 2008 ((Republic of The Sudan and the Sudan People's Liberation Movement/Sudan People's Liberation Army, 2005:17). The result of this new view of statistics was that more government resources were being allocated to NSS. This led to improved statistical legislation and increased data production on the continent during the 1990s (Lehohla, 2008:10).

## A DATA CURATION HISTORY OF AFRICA

However, there were still a number of challenges for data management on the continent. These included the need for more statistical advocacy to promote the value of statistics to policy-makers, and strategies to improve the quality of African data, including the development and promotion of data curation standards (United Nations Economic and Social Council, Economic Commission for Africa, 2001:8). Often international efforts during this decade focused on statistical development in a particular African region. For example, in 1997 EUROSTAT, the statistical office of the European Commission (EC) launched a programme for statistical development in the West African Economic and Monetary Union (WAEMU). However, these regional programmes were uncoordinated and thus failed to increase data capacities generally in Africa (Lancetti, 2004:12). African statistical development programmes in the 1990s were still limited by a strong focus on data collection, rather than on the overall curation of African data. International statistical programmes linked data collection to national development but failed to promote all aspects of data management on the continent (Onsembe & Hie, 2004:10).

At the end of this decade African data management capacities were aided by the introduction of a new results-based development framework supported by the international donor community. The International Monetary Fund's (IMF) Poverty Reduction Strategy Programme, introduced in 1999, required extensive national data from member countries to allow the IMF to monitor the programme's efficacy. IMF borrower countries were encouraged by donor bodies to draw up Poverty Reduction Strategy Papers (PRSP) indicating their plans for future growth, and these were required to include statistical improvement plans to provide the data for the monitoring of country progress, for both the development agency and the partner country (Dupriez, 2008a:2). The renewed demand for reliable statistics initiated by the PRSPs provided the stimulus for donor organisations to increase their efforts to support all aspects of data curation in developing countries. This provided African governments with an opportunity to access donor funding and expertise to advance the data management aspects of their statistical systems (Achikbache, et al., 2002:159).

A key collaborative grouping established in the 1990s for data development was the Partnership in Statistics for Development in the 21<sup>st</sup> Century (PARIS21). The PARIS21 Consortium constituted policymakers, policy analysts and statisticians and aimed to foster evidence-based decision-making, particularly in developing countries, through improving the production and use of statistics in these countries. The work of the consortium was sponsored by the EU, IMF, OECD, UN and World Bank (Kiregyera, 2001:2). The formation of strategic partnerships such as this to

co-ordinate efforts at effective information utilisation in Africa accounted to some extent for the progress in the following decade with regard to data curation programmes in the region.

### **5.4.2 Development of African Institutional Infrastructures in the 1990s**

Development frameworks of the 1990s emphasised the need to advance the regulatory aspects of information management in UNECA member states, as well as the human and technological resources required for the management of information for social and economic development. These plans emphasised the need to establish mechanisms for the continuous collection and management of African data (United Nations Economic and Social Council, Economic Commission for Africa, 1999: Infrastructure Development (Information Resources, Content and Applications). This was translated in this period into renewed efforts to strengthen the capacities of African NSOs. For example, UNECA's new "harnessing information for development" programme addressed the need to invigorate statistical systems in African countries through the provision of funding for and technical assistance to African NSOs for data collection and data management. NSOs were seen as key institutions in the statistical development process, because most of the human and technological capacities for data curation in Africa resided in these agencies.

#### **5.4.2.1 National Data Producers and Support for Data Quality**

The AAPA delegated the responsibility for co-ordinating national statistical activities to NSOs in each country, while UNECA was tasked with regional co-ordination of these activities (Ching'anda, 1999:1). Some of the AAPA's recommendations with regard to strengthening NSOs in Africa were put into effect during the 1990s. For example, in 1993 the governments of Francophone African countries established the Economic and Statistical Observatory for Sub-Saharan Africa (AFRISTAT) to support the development of NSOs in member states of Francophone Africa (Economic and Statistical Observatory of Sub-Saharan Africa, 2009).

Another regional initiative in this regard was the Mediterranean Statistical Co-operation programme (MEDSTAT). EUROSTAT launched this programme in 1996, aimed at statistical cooperation between the EU and Mediterranean countries, including the North African states of Egypt, Morocco and Tunisia. The programme fostered the supply of reliable data to assist the political and economic agreements between the EU and partner countries, and to contribute to

## A DATA CURATION HISTORY OF AFRICA

sound government planning in these countries. The first phase of the programme ran from 1996 to 2003 with a budget of 20 million Euros (European Commission, 2009c).

Developments in the 1990s included an emphasis on the *quality* of statistics produced in the region. This was inevitably linked to support for international standards in the production, management and dissemination of data. Development planning on the continent is dependent on research based on reliable data, and donor organisations became aware that supporting the quality dimensions of data production and usage was an important component of statistical capacity building in the region. Research conducted in the early 1990s showed problems with reliability and consistency in the data from surveys conducted in Africa. A World Bank sponsored conference of the African Economic Research Consortium (AERC), held in June 1991, dealt with serious inconsistencies in the findings of surveys conducted by different organisations producing data on the economic indices even of the same country (Ademola & Adeniran, 1998:2-3). Researchers and donor organisations were concerned that diverse findings based on this inconsistent data could impair African governments' ability to make sound policy decisions.

In response the AERC commissioned a study in the early 1990s of the data collection procedures used by international organisations conducting surveys in Africa in collaboration with African NSOs. The study confirmed the existence of inconsistencies and discrepancies in survey data from these organisations, noting that even intra-agency publications sometimes contained conflicting data (Ademola & Adeniran, 1998:3). Recommendations of the study included the strengthening of survey research capacity in Africa, and the use of internationally accepted standards with regard to concepts and definitions by those organisations conducting survey research in Africa (Ademola & Adeniran, 1998:34).

Thus donor organisations introduced data quality frameworks to boost the quality of African data. Data accessibility was one of the data quality dimensions targeted by the international community during this period. This occurred in the wake of the international financial crisis of 1994-1995, which originated to some extent from a lack of consistent and reliable economic information among trading countries. In 1996 the IMF launched the Special Data Dissemination Standard (SDDS) as a framework for improving statistics of countries to aid their access to international markets (Achikbache, et al., 2002:472). Their General Data Dissemination Standard (GDDS) was introduced the following year as an international standard for data production for countries unable to maintain the statistical standards required of the SDDS. The

GDDS provides a framework for participating countries to assess their NSS and to receive assistance in strengthening the quality of the data they produce (International Monetary Fund, 2009).

### **5.4.2.2 Other Data Curation Institutions in Africa in the 1990s**

Research support institutions to assist data users play a complementary role to NSOs in Europe and North America. However, very few such organisations exist in Africa. The first efforts to support institutional frameworks for data management beyond NSO development occurred in 1993 with the establishment of the South African Data Archive (SADA) in Pretoria, South Africa. SADA was set up at the Human Sciences Research Council (HSRC) and was based on the European SDA model. It is now a unit of the National Research Foundation, a research support facility in South Africa (Lesaoana, 1997:4; South African Data Archive [SADA], 2009). This was the only such development in Africa, however, as survey data production and archiving in other African countries continued to be almost exclusively the task of NSOs.

### **5.4.3 ICT Infrastructures for Knowledge Management in Africa in the 1990s**

During the 1990s coordinated efforts were initiated to build African ICT infrastructures to take advantage of new technologies for effective data curation in the region. In 1992 the UNECA Conference of Ministers passed a resolution (732, XVII) calling for member states to adopt ICT policies for the effective utilisation of information resources (United Nations Economic and Social Council, 2001:8). They had the sanction of the international community in this endeavour, particularly after 1995 when the G7 Conference<sup>4</sup> in that year supported requests from African leaders to assist them to position themselves more favourably within the global information society (Nassimbeni, 1998:158).

As part of its strategy to aid the development of ICT infrastructures in countries of the region UNECA, through its PADIS programme, organized the first African Regional Symposium on Telematics for Development in Addis Ababa, Ethiopia in 1995. From this symposium arose the African Information Society Initiative (AISI) launched in 1996. This plan became the strategic

---

<sup>4</sup> Canada, France, Germany, Italy, Japan, the United Kingdom and the United States were members of the G7 grouping.

framework for information development in Africa, including ICT infrastructural advancement to enable the utilisation of information technologies for effective information management in African countries (United Nations Economic and Social Council, Economic Commission for Africa, 1999:2, 9; James, 2001:169). The AISI received the sanction of the African Telecommunications Ministers at an African Regional Telecommunications Development Conference held in Abidjan, in May 1996 and was endorsed by the Organisation of African Unity (OAU) Heads of State Summit held in Cameroon in July that same year (United Nations Economic and Social Council, Economic Commission for Africa, 2008).

The AISI placed emphasis on raising awareness of the value of ICTs for development, furthering national ICT infrastructure policies, and promoting internet connectivity in African countries. These strategies were aimed at ensuring that African countries were able to participate in the information age to their benefit. Furthering the development of ICT infrastructures for the effective utilisation of development information does not translate automatically into support for data curation on the continent. However, national digital management plans cannot be put in place in African countries without the existence of ICT infrastructures for their realisation.

A Committee on Development Information (CODI) was established by UNECA at this time to implement the AISI (United Nations Economic and Social Council, 1997:1). Working within the AISI framework, UNECA encouraged African governments to create policies to build the physical and regulatory infrastructures necessary to take advantage of developments in technology for the effective harvesting of their information resources. These policies on National Information and Communication Infrastructure (NICI) are steps African governments undertake to develop their information technology infrastructures in order to harness ICTs for the attainment of development objectives (e-Strategies, 2007:1-6).

### **5.4.4 Human Resource Development for Data Curation in the 1990s**

The AAPA assessment emphasised the need for improvements in regional data sharing to advance the information society in Africa and support sound governance in countries of the region. The study also advocated the introduction of new statistical training programmes in the region to replace the UNDP funded training programmes which ceased in 1993.

The establishment of more African universities during this decade increased human resource capabilities in the region to assist data management. By the end of the 1990s formal statistical

training was being conducted at regional STCs in Cameroon, Côte d'Ivoire, Morocco, Tanzania and Uganda, and at the University of Ibadan in Nigeria. Informal training at professional and sub-professional levels took the form of on-the-job training, as well as infrequent seminars and workshops conducted by African STCs and NSOs. Formal training at STCs was hampered by staff and skills shortages. Formal training programmes were also hampered by weak linkages between STCs and NSOs in African countries, leading to their lack of relevance to the needs of African NSS (United Nations Economic and Social Council, Economic Commission for Africa, 2001:15-16).

Statistical training programmes in this decade did not include specific data management training. However, in South Africa SADA promoted skills development in this regard through workshops and meetings. For example in 1998 a SADA workshop on "Data sharing for research capacity development" drew 150 participants from government, Non-Governmental Organisations and research institutions. Government funding and institutional support from the National Research Foundation played a key role in enabling the data management training initiatives of SADA (South African Data Archive, 1998:3).

### **5.4.5 Professional Associations**

Training in data curation needs to be bolstered by robust professional associations to encourage mentoring and the exchange of expertise among practitioners. During the 1990s there were few professional associations for African statisticians and none at all for African data managers. The statistical associations that did exist were shown to be poorly supported by African governments (United Nations Economic and Social Council, Economic Commission for Africa, 2001:10-13).

## **5.5 Data Curation in Africa from 2000**

The advent of the new Millennium and the international donor communities' formulation of the Millennium Development Goals (MDGs) as a development framework for the 2000s highlighted the need for country data for the monitoring of national and international development plans. As with PRSPs, the MDGs' new results-based agenda gave a higher profile to statistical information, as key input to provide indicators for the efficacy of international development programmes (Lufumpa & Mouyelo-Katoula, 2005:30). Through the creation of their MDG strategies, African governments committed themselves to achieving measurable goals in poverty

reduction. However, their statistical systems were shown to be inadequate to meet data requirements for planning for and monitoring the achievement of these goals (Simonpietri & Williams, 2005:14).

### **5.5.1 Coordination of International Support for Data Curation in Africa from 2000**

African governments faced several impediments to advancing their NSS for the production of development data. Consequently during this period a number of new African statistical capacity-building initiatives were undertaken by regional and international donor organisations to advance statistical capacities in the region.

#### ***The Coordinating Committee on African Statistical Development (CASD) Review***

In 2001 UNECA's Coordinating Committee on African Statistical Development (CASD) undertook a detailed assessment of the data situation in Africa, in relation to the advances in this sphere initiated by the AAPA. CASD examined the state of NSS in African countries, investigated the quality of data emanating from African institutions, and evaluated donor programmes aimed at advancing the production and use of statistics on the continent. The study also investigated statistical training programmes in the region and the role of regional and sub-regional organisations in statistical capacity building. The CASD review showed that, while some of the recommendations of the AAPA were heeded and improvements in African NSS were made during the 1990s, on the whole the plan was not sufficiently publicised or popularised, and was hampered by resource shortages within African NSS due to lack of support from African governments (United Nations Economic and Social Council, Economic Commission for Africa, 2001:10).

Spurred by the CASD assessment, donor organisations initiated joint projects for data development on the continent. In 2003 the African Capacity-Building Foundation (ACBF), a development body formed by African governments in partnership with the World Bank, the African Development Bank, and the United Nations Development Programme, established the Statistics Technical Advisory Panel and Network (STATNET) for the sharing of best practice in statistical production in the region. The ACBF sponsored other data development programmes in Africa during this decade. This organisation funded the International Comparison Program for Africa (ICP-Africa), a statistical capacity building program for fifty two African countries. ICP-Africa was run under the ICP Program, an international project to boost country capacities for

generating Purchasing Power Parity estimates for cross-country price comparisons. Unlike ICP-programs in other parts of the world, the ICP-Africa program had the expressed aim of strengthening African NSS (African Capacity Building Foundation, 2004:3; Lufumpa & Mouyelo-Katoula, 2005:30-31).

### ***The Marrakesh Action Plan***

In 2002 a United Nations Conference on “Financing for development results” held in Monterrey, Mexico produced the Monterrey Consensus and a UN conference in 2005 in Paris led to the Paris Declaration on Aid Effectiveness. These initiatives were concerned with streamlining international donor support and encouraging donor-country partnerships for better aid spending. They emphasised the need for good policies and good institutions as requirements for effective development programmes (Regional Reference Strategic Framework for Statistical Capacity Building in Africa [RRSF], 2006:131-132).

These initiatives to achieve better coordination of support for developing countries culminated in a World Bank organised conference held in February 2004 in Morocco on “Managing for Development Results”. This conference investigated international areas of cooperation in managing and monitoring indicators of development at country level and produced a global action plan for the advancement of statistical systems to monitor development outcomes, the Marrakesh Action Plan for Statistics (MAPS). MAPS committed the international community to endorsing data curation for growth and good governance in African countries (and other developing countries). It was formulated from lessons learned by international donor agencies through previous statistical capacity building projects in these countries (Marrakesh Action Plan for Statistics, 2004). MAPS was endorsed by the UN Statistical Commission (UNSTATS), representing the international statistical community, the ADB, OECD, PARIS21 and UNECA. It built on strategies initiated by the 1990 AAPA, and on the GDDS's success at gaining the participation of African countries (Eele & Chinganya, 2005:23).

MAPS encouraged the mainstreaming of statistical systems in country development plans such as PRSPs, and increased funding for statistical capacity-building in developing countries. To achieve these goals the plan proposed that all low-income countries formulate National Strategies for the Development of Statistics (NSDS) by 2006, in time for a proposed international MDG review in 2010. NSDS are frameworks for the development of countries' statistical systems, based on international best practice. The PARIS21 Consortium was given a mandate by the international community to focus on the creation of NSDS in African countries

and has produced guidelines for the formulation of these strategies. Funding for this was provided by the World Bank's Trust Fund for Statistical Capacity Building and other international donors (Partnership in Statistics for Development in the 21st Century [PARIS21], 2008:5; Partnership for Statistics in the Twenty-first Century [PARIS21], 2009).

### ***The Forum for African Statistical Development***

The Forum for African Statistical Development (FASDEV) was launched in 2004, sponsored by the ADB, PARIS21, UNECA and the World Bank. The aim of this series of annual meetings is to coordinate data development projects in the region and reduce duplication of efforts in this regard. The FASDEV I meeting assessed existing African statistical development frameworks such as MAPS, PRSPs, and regional strategic initiatives, and drafted a planning document, the Regional Reference Strategic Framework (RRSF), as a guide for statistical capacity building in the region (Lufumpa & Mouyelo-Katoula, 2005:38-39). This new international plan to boost African data production capacities was launched in 2006 at FASDEV II and received funding of US\$30 million from the African Development Bank and international development partners (United Nations Economic and Social Council, Economic Commission for Africa, Forum on African Statistical Development [FASDEV], 2006).

The RRSF continues the focus on ensuring that sustainable international statistical development programmes are put in place in the region. It advocates for better management of NSS in line with the UN Fundamental Principles of Statistics. Like earlier such initiatives, it supports the advancement of ICT infrastructures in African countries for data curation and human resource development to strengthen NSS in the region, particularly the development of data analysis skills among African researchers. The ADB and UNECA were tasked with coordinating the implementation of the RRSF. NSOs were expected to be the main drivers of the RRSF within African NSS, with African governments facilitating the process, and providing funding alongside donor contributions (Regional Reference Strategic Framework for Statistical Capacity Building in Africa [RRSF], 2006:133-134).

### ***The African Symposium on Statistical Development***

The UN initiated African Symposium on Statistical Development (ASSD) is an important example of international and regional support for official data collection and dissemination on the African continent during the 2000s. These annual symposia were convened in 2006 as a platform for the exchange of information and technical assistance concerning African censuses. At the inaugural symposium delegates, representing statistical agencies in forty three African

countries, pledged to promote “knowledge management in statistics” on the continent (Africa Symposium on Statistical Development, 2006a:4).

The ASSD assists with data advocacy in the region. In 2007 Africa’s NSOs, in collaboration with the African Development Bank, the African Union Commission and UNECA, created an African Charter on Statistics which acknowledged the vital role quantitative data plays in development on the continent. The charter was adopted by the Assembly of Heads of State of the African Union in February 2009 and is an important step toward formalising country support for data curation on the continent (African Union, 2009).

### **5.5.2 Regional Initiatives**

Collaboration among African statistical organisations for the production of good quality survey statistics was boosted by the establishment of the African Union (AU) in 2002 which promotes regional development and the development of an African economic community (African Union, 2006?). The AU is taking active steps to promote statistical growth through the establishment of funds and the creation of appropriate statistical policies to support regional statistical advancement (African Union, Economic Affairs Department, Statistics Division, 2008: sections 1-2). UNECA continued its coordination role in regional data advancement projects and in 2006 UNECA’s African Centre for Statistics (ACS) was created to focus on the organisation’s role in promoting knowledge management in the region (Partnership in Statistics for Development in the 21st Century [PARIS21], 2009; United Nations Economic and Social Council, Economic Commission for Africa, 2008). In 2006 the West African statistical coordinating agency, AFRISTAT, launched its Strategic Plan of Activities (ASPA) 2006-2010, which acknowledged the historical neglect of data dissemination strategies in statistical capacity-building programmes, and included plans to rectify this (Balepa, 2006:135-6).

Regional statistical capacity building initiatives included the programmes of the European Union’s statistics office, EUROSTAT (Lancetti, 2004:12). North African countries continued to receive international assistance through the MEDSTAT II Programme, launched in 2006 with a budget of 30 million Euros (European Commission. EUROSTAT, 2009c). The IMF and World Bank established programmes in Anglophone and Francophone Africa, with the assistance of regional support structures such as UNECA and AFRISTAT. Regional cooperation among NSOs included the sharing of skills and experiences among these institutions. For example as part of the IMF’s Anglophone Africa project statisticians from Zambia shared technical skills with the

Botswana Bureau of Statistics for the processing of their trade data (Eele & Chinganya, 2005:24-25).

### **5.5.3 Development of Regulatory Infrastructures for Data Curation in Africa: 2000 Onwards**

The CASD assessment revealed African NSS to be governed by outdated legislation which did not ensure the autonomy of their NSOs and therefore the integrity of their data products. The decentralised nature of African NSS was also seen to hamper sound data quality as this led to uncoordinated data production, with the collection, archiving and dissemination of statistics carried out by a number of government ministries and departments. NSOs were shown to be too under-resourced to fulfil their legally mandated coordinating role within the NSS. NSS were also shown to be unsuitably structured to meet data users' needs. The establishment of user/producer committees, recommended by the AAPA, was only implemented in some countries, and did little to change the knowledge utilisation culture of these countries (United Nations Economic and Social Council, Economic Commission for Africa, 2001:10-11).

The earlier AAPA evaluation had recommended advocacy to key stakeholders in African statistical systems regarding the value of data as input to governance (Simonpietri & Williams, 2005:12). The CASD study supported this to counter the lack of statistical awareness among researchers and the general population in African countries. African Statistics Day, introduced by the AAPA as an advocacy strategy, was shown to be celebrated by some African countries, but with little impact. African governments too, failed to promote the use of statistical information for planning and research and development strategies (United Nations Economic and Social Council, Economic Commission for Africa, 2001:12-14).

The 2000s saw a gradual change of attitude among African leaders towards information management. The historically suspicious attitudes of African governments towards academic research gave way somewhat to an understanding among African leaders of how empirical research can inform national decision-making and lead to more effective and legitimate government action. However, commitment to data sharing for economic and social development has not been evidenced in government policymaking in the region. Thirty four African countries published Poverty Reduction Strategy Papers between 2000 and 2005 and twenty eight of these papers, which are published on the IMF website, acknowledged the importance of empirical data for development planning. However, these contained only

passing references to concrete data development plans (International Monetary Fund [IMF], 2009b).

As discussed earlier, a suggestion of the AAPA to counter government apathy in this regard was the creation of national statistical development plans integrated into country development agendas (Simonpietri & Williams, 2005:12-13). The PARIS21 group has been tasked with furthering the creation of these NSDS by African governments. Currently the governments of twenty seven of the fifty three African countries are implementing NSDS (Partnership in Statistics for Development in the 21st Century [PARIS21], 2009).

### **5.5.4 Institutional Infrastructures: 2000 Onwards**

Despite the existence of international programmes to boost statistical capacities in countries of the region, African NSS were still weak at the start of the millennium. These systems were undeveloped, with sparse provision for assessing the data needs of users. African governments showed little commitment to utilising empirical data for decision-making and economic growth, and support for the development of NSOs had not been mainstreamed into their national policies.

#### **5.5.4.1 National Statistics Offices and Data Quality Initiatives from 2000**

A 2002 African Development Bank assessment of the statistical capacities of forty-nine African countries revealed that their NSOs were unable to provide reliable socio-economic indicators for comparative purposes owing to structural, financial and human resource constraints (Lufumpa & Mouyelo-Katoula, 2005:30-31). This was despite the fact that by then all but eight African countries were subscribing to the IMF's SDDS and GDDS data quality frameworks (International Monetary Fund [IMF], [2006?]). To improve this situation the IMF introduced a new quality framework in 2001, the Data Quality Assessment Framework (DQAF). The DQAF operates alongside the SDDS and GDDS but expands on these frameworks by using five attributes to judge data quality: integrity, methodological soundness, accuracy and reliability, serviceability, and accessibility. It acts as a best practice guide for the IMF to formulate assessments of the quality of country data.

At present the value of this framework lies in the information provided by country Reports on the Observance of Standards and Codes (ROSCs), which are compiled using the framework

## A DATA CURATION HISTORY OF AFRICA

and posted on the Dissemination Standards Bulletin Board website of the IMF. ROSCs present an overview of institutional arrangements of NSS, their data collection, archiving and dissemination processes and the usability of their statistical products. To date eighteen African countries have undergone these assessments, which contribute to the knowledge-pool to assist in identifying the development needs of these institutions (International Monetary Fund [IMF], 2009a).

Some African NSOs have put in-house systems in place for quality control of survey datasets. For example, Statistics South Africa (StatsSA) has developed a central database known as the Data Management and Information Delivery (DMID) system which supports the use of standard data items through the production of standard metadata for its surveys. The DMID provides the South African NSO with the tools to evaluate the quality of their statistical data and assists the production of accurate and comparable national statistics (Lehohla, 2007b). The DQAF aims to assist NSOs in other African countries to develop systems of this nature.

In the current decade most NSOs in Africa still do not have effective systems in place for the archiving and sharing of their microdata. These systems are needed to ensure efficient data discovery and dissemination and the standardization of metadata for data comparability. To support this, international donor organisations have begun to focus on the provision of software and guides for data management in the region. Examples are the data management tools developed by the International Household Survey Network (IHSN). The IHSN was established in 2004 as a recommendation of MAPS. The Network aims to improve the quality of survey data in African countries (and other developing countries) to promote data usage for research and policymaking in these countries. Its membership is comprised of organisations that provide funding and technical support for household survey programmes and includes, among others, PARIS21, UNSTATS and the World Bank. The IHSN has also undertaken a data audit for African countries, creating a central data catalogue accessible via the IHSN website (International Household Survey Network, 2006).

The tools developed by the IHSN deal with the full survey life cycle, from sampling and questionnaire design to data archiving and dissemination. One of these tools is the National Data Archive (NADA) toolkit, which is an open-source web application that allows the creation of searchable online data catalogues and microdata dissemination via the internet. The IHSN's Accelerated Data Program (ADP), implemented as a satellite program of PARIS21, provides training and assistance in data curation to NSOs in developing countries, utilising the tools

designed by the IHSN (Dupriez, 2008b; Greenwell, 2008; International Household Survey Network, [2008?]). In keeping with international strategies to form partnerships for effective capacity building, the IHSN/ADP Project initiated a partnership with DataFirst to undertake NADA software installations and training in African NSOs from 2008.<sup>5</sup>

To date the NADA system has been installed at NSOs in eleven African countries - Ethiopia, Gambia, Ghana, Lesotho, Liberia, Mozambique, Niger, Nigeria, Senegal, Sierra Leone and Uganda. The Nigerian Bureau of Statistics (NBS) is the second African NSO (after Statistics South Africa) to provide trouble free public access to national microdata via their website, and this has been accomplished through the installation of the NADA application at this NSO. Requests for the Nigerian datasets have been received from researchers abroad and in Africa, providing evidence that data supply creates demand even in a region where microdata use is not extensive. Through NADA installations and training the IHSN/ADP project is advancing data curation on the continent, and supporting data quality improvements in the region. However, skills shortages and technical issues are hampering optimal application of the software to promote data accessibility and data sharing in African countries participating in the project.

### **5.5.4.2 Survey Data Archives in Africa**

In 2000 DataFirst, a second African SDA, was established at the University of Cape Town in South Africa, with donor funding. Unlike SADA, which is primarily a data preservation institution, DataFirst was created to promote quantitative skills among social science researchers in the country. This was later extended to include an initiative to monitor the quality of official South African survey data (DataFirst, 2009b). DataFirst and SADA were established because International grant-makers believe that science and policymaking will benefit from strong research support structures to promote the sharing of survey data.

---

<sup>5</sup> In 2008 the author participated in NADA installations and training at the Lesotho Bureau of Statistics and the Mozambican Institute of National Statistics.

### **5.5.5 Support for African Technological Infrastructures for Data Curation in the 2000s**

From 2000 ICT strategies were formalised by many African countries, with the support of UNECA and its development partners. New international programmes aiding information utilisation for development were introduced at this time. For example, in 2001 UNESCO launched its Information for All Programme (IFAP) which was an intergovernmental programme concerned with global equitable access to information resources. The IFAP encouraged countries to harness the opportunities provided by the information society for their social and economic development (United Nations Educational, Scientific and Cultural Organization (UNESCO), 2009).

Utilising partnerships with international development agencies, UNECA assists African countries in the building of their information technology infrastructures. By 2006 thirty African countries had NICI plans in place (United Nations Economic and Social Council, Economic Commission for Africa, 2007:4). That year UNECA initiated the Regional Information and Communication Infrastructure (RICI) to act as a co-ordinating framework for the NICIs of the Regional Economic Communities in Africa. The aim of the RICI is to develop harmonised regional ICT frameworks, and this has included the creation of draft guidelines for data confidentiality in West Africa and the Central African States, where no legal protection for this exists at present, and the development of e-commerce strategies for the Arab Maghreb states of Libya, Morocco and Tunisia (United Nations Economic and Social Council, Economic Commission for Africa, 2007:41-48).

Other programmes related to ICT development in Africa were launched in the second half of the decade. A Connect Africa Summit was held in Kigali, Rwanda, in October 2007 at which forty three African countries were represented along with African development banks and international development agencies. The Summit ended with delegates' adoption of five goals to bridge the digital divide in Africa, including the establishment of healthy ICT infrastructures in African countries. The summit provided support for the implementation in Africa of the connectivity goals of the 2003-2005 World Summit on the Information Society (WSIS) which was concerned with the development of ICTs to support the achievement of the MDGs (World Summit on the Information Society, 2003:1-2).

### **5.5.6 Human Resource Development for Data Curation from 2000**

International donor organisations and regional support structures continued to foster training in the region for statistical development, and in this decade specific training in data management began to be included in training curricula, albeit to a limited extent. In 2002, after requests by the New Partnership for Africa's Development (NEPAD), the IMF established African Technical Assistance Centres (AFRITACs) in the region to provide training to strengthen African countries' ability to produce relevant statistics for the development and monitoring of their macroeconomic policies and plans to feed into their PRSPs. The East African AFRITAC was located in Dar es Salaam, Tanzania, and the West African AFRITAC was established in Bamako, Mali (International Monetary Fund, East Africa Regional Technical Assistance Centre (AFRITAC), 2004).

At this time the ADB also began working with the six regional STCs in Africa to modify their curricula for the ICP-Africa program. This included co-ordination of regional training and the development of training materials (Lufumpa & Mouyelo-Katoula, 2005:43). International training assistance with statistics in the region continued with the UNFPA'S introduction in 2002 of its Technical Advisory Programme (TAP). The programme's Technical Advisory System has three Country Technical Services Teams (CSTs) in Sub-Saharan Africa - in Addis Ababa, Dakar and Harare. It has forty five country programmes and forty three field offices in the region. The programme was extended in 2006 as the Technical Assistance Programme (TAP). Problems with training programmes during this time included the overly-theoretical nature of the training offered at the STCs, and staff shortages in these institutions. Weak linkages between STCs and NSOs in African countries led to training programmes at these centres that were not always relevant to the needs of African NSS (United Nations Economic and Social Council, Economic Commission for Africa, 2001:15-16).

These training initiatives also currently focus on data collection, data cleaning and data analysis techniques, with little emphasis on data preservation for secondary use. The paucity of training in data management and data archiving in the region hampers the development of skills in this field. Training in data preservation and dissemination techniques is generally presented as short courses or once-off workshops (United Nations Economic and Social Council, 2001:16). For example, the ISAE in Uganda ran a short course on data management in 2001 (Nyanzi, 2004:6). EASTC in Tanzania runs a certificate course for Statistical Assistants in data

collection and data management procedures for the production of official statistics (Eastern Africa Statistical Training Centre [EASTC], 2006; Muba, 2006:148-9). The Cairo Demographic Center, an Egyptian government agency, also offers short courses in data (Cairo Demographic Center, [2009?]). The IMF's recent data quality initiatives have included a data management component undertaken as part of their regional GDDS Technical Assistance project (International Monetary Fund [IMF], 2009a).

Some in-service training related to data management takes place at those NSOs which have the necessary facilities, such as the Nigeria Bureau of Statistics, which has three training schools for the training of "sub-professional statisticians" in various aspects of the data collection and data management (Akinyosoye, 2008: 213-214). Training in the use of IHSN tools is generally undertaken by the IHSN/ADP at NSOs but some training in these tools is carried out by staff in-house, for example at the Central Statistical Authority (CSA) of Ethiopia (Central Statistical Authority of Ethiopia (CSA), 2008). Statistics South Africa also conducts training and advocacy workshops on data management, such as an African Statistics Day focus group workshop held in 2006.<sup>6</sup>

Training courses in survey data analysis are also sparse in Africa. One example is the work undertaken by DataFirst at the University of Cape Town (UCT). This archive has been tasked with building quantitative skills among social science graduates and faculty at UCT and co-facilitates an annual survey data analysis course in collaboration with academics from UCT and the University of Michigan in the USA (DataFirst, 2009c). Another example, also from South Africa, is the survey design and analysis course offered at the School of Development Studies at the University of KwaZulu-Natal (UKZ-N) (Bailey, 2005:540).

### **5.5.7 Professional Associations in the Region: the Establishment of the African Association of Statistical Data Archivists<sup>7</sup>**

The absence of national and regional associations of data managers hampers human resource development in this field in Africa (United Nations Economic and Social Council, Economic Commission for Africa, 2001). Professional associations allow for communication among

---

<sup>6</sup> The author attended this and other StatsSA workshops held in Cape Town.

<sup>7</sup> This section is an abbreviated version of "The establishment of the African Association of Statistical Data Archivists" by Lynn Woolfrey published in *IASSIST Quarterly* 31(2):24-26, 2007.

professionals in a field of practice (Mouyelo-Katoula, 2006: 29). Such communities of practice provide a supportive network for mentoring and other forms of skills exchange. They can also act as pressure groups to effect changes in society, in this case to encourage competent management of national data for the benefit of society. Such a community was initiated when African delegates at IASSIST's 2007 conference formed an interest group concerned with data management in Africa. The IHSN/ADP project provided funding and technical assistance for this group to establish an association.

The African Association of Statistical Data Archivists (AASDA) was inaugurated in 2008. The initial meeting represented the first occasion where data managers from Francophone, Lusophone and Anglophone African countries had met. The Association is mandated to work towards overcoming organisational issues hindering data production and data sharing in the region. AASDA aims to advocate for the creation of good metadata for African datasets, based on international standards, and for the incorporation of best practice in the preservation and dissemination of data as part of their organisational objectives (African Association of Statistical Data Archivists, 2008). Members receiving training in the use of data curation tools can play a mentoring role on the continent, nurturing skills nationally and regionally.<sup>8</sup>

The creation of AASDA can be seen as a stage in the realization of data sharing on the continent through promoting linkages among survey data managers in the region. The Association will be in a position to endorse the establishment of data management facilities on the continent. The formation of a group of dedicated practitioners in the field can serve to highlight the need for coordinated microdata curation in Africa. The initiation of a network of survey data managers in Africa represents the first step towards such collaboration.

### **5.5.8 The 2000s: The Sharing of Publicly Funded Research Data**

Support for the production, sound management and effective use of official statistics in Africa has been complemented by a parallel movement promoting the management and open exchange of digital information from publicly funded research. This Open Data movement is part of the "Open" ethos – including the Open Access, Open Archives and Open Source Software movements. These initiatives promote the view that technologies or research findings

---

<sup>8</sup> The author represented her organisation, DataFirst, at AASDA's inaugural meeting in Cape Town in 2008.

## A DATA CURATION HISTORY OF AFRICA

should be made freely available for research and development to enable them to become a collaborative resource for the creation of a more equitable society. The Open Data movement refers to social science data, technical information and government statistics, and is not limited to survey microdata, but includes this form of data.

This movement supports wider access to research results to maximize their value for the research and policy communities, and society in general. The underlying assumption is that countries benefit socially and economically by treating information as an inexhaustible resource that in fact gains value with use and that sharing data from publicly funded research leads to better returns on this investment (Organisation for Economic Cooperation and Development [OECD], 2007:9). The philosophy of this movement is also based on a belief that the current unequal access to information globally limits the advantages of research and therefore wastes public resources. This is especially true of survey resources, as data from surveys conducted internationally, even those using African data for comparative purposes, is often not freely available to researchers within African countries. African researchers' potentially valuable local input is also not available to the international community because African scholars are not able to access all African data for secondary analysis.

In 2000 the International Council for Sciences' CODATA Committee (Committee on Data for Science and Technology) established a CODATA Task Group on Preservation of and Access to Scientific and Technical Data in Developing Countries (the task group involves Brazil, China and South Africa.). Although CODATA and its task groups are concerned with scientific and technical data, their work has policy implications for the sharing of data in the social sciences (International Council for Science, Committee on Data for Science and Technology [CODATA], 2009). CODATA's focus is on identifying existing datasets and practitioners in the region and on training for data managers, which are both also important aspects of social science data management (Page-Shipp, 2007:23). The sharing of publicly funded research data has also been supported by international initiatives aimed at promoting the information society. For example, the 2003 WSIS produced a declaration on "information generation and knowledge development" advocating "open access to primary scientific data and publications" and calling on public institutions to "promote the results of activities which have been funded by public money" (World Summit on the Information Society, 2003:21).

These initiatives towards the sharing of government sponsored research culminated in countries adopting the OECD's Declaration on Access to Research Data from Public Funding, in 2004

## A DATA CURATION HISTORY OF AFRICA

(Organisation for Economic Cooperation and Development [OECD], 2007:7). South Africa was the only African signatory to this document. The declaration focuses mainly on data produced in scientific and technological research but encompasses all research data. It is therefore relevant to the sharing of social science data, including empirical data from social surveys. The declaration placed data sharing on the official agendas of several governments, including the South African government. This international commitment required the South African government to make policy changes with regard to data curation in the country. The South African Department of Science and Technology (DST) was tasked with taking the initiative forward by creating government policy with regard to data access.

The DST's work in this regard led to a workshop on National Access to Research Data, in Pretoria in 2007. The DST worked to establish a national Network of African Data and Information Curation Centres (NeDICC) to provide the national infrastructure for data collection and data sharing and to form partnerships and facilitate training for data curation in South Africa (Gray, 2007:1). In 2008 the DST organised the First African Digital Curation Conference on February 12th-13th in Pretoria, South Africa. At this conference, stakeholders in data preservation and data sharing from the physical and social sciences met to provide input to government policy on the sharing of digital data from research<sup>9</sup>. Their mandate related to research data from many disciplines but policy outcomes from this development would be relevant for social survey data sharing in South Africa and other African countries.

Unfortunately the loss of key personnel at the DST who were involved in the project seems to have led to this initiative acquiring a narrower focus. Resulting legislation reflects government concerns around the patenting of potentially lucrative research findings. The DST initiated "Intellectual Property Rights from Publicly Financed Research and Development Act" enacted in 2008 deals with maximising national gains from publicly funded research and development with regard to scientific and technological innovations (Republic of South Africa, 2008:4). The definitions summarised in Section 1(c) of the Act (Republic of South Africa, 2008:3) are deficient in that they do not address the sharing of the findings from social science research, and in fact specifically exclude copyrighted academic work (Republic of South Africa, 2008:3). The 2009 conference on African Digital Scholarship and Curation was a more low-key affair than the first conference, with no government sponsorship of delegates (University of Pretoria and University of Botswana, 2009). Legislation to promote the sharing of government sponsored

---

<sup>9</sup> The author gave a presentation at this conference

## A DATA CURATION HISTORY OF AFRICA

social science research data has yet to be promulgated by the South African government (Gray, 2009; Page-Shipp, 2008).<sup>10</sup> Similar open data movements have not been established in other African countries.

This chapter provided an historical overview of data curation in Africa, including seminal events which promoted the use of survey and census microdata for research and policy formulation on the continent. The next chapter examines the nature and extent of data usage by African governments to assist national planning.

University of Cape Town

---

<sup>10</sup> At the time of this research the DST had not produced any documentation from the workshop. Information on the workshop was obtained from Gray (2007) and discussions with Daisy Selematsela of SADA).



Photograph 7: The Central Agency for Public Mobilization and Statistics (CAPMAS) in Egypt (Photo: CAPMAS website, 2008: <http://www.capmas.gov.eg>).



Photograph 8: The Gambia Bureau of Statistics (GBOS). (Photo: GBOS website, 2009: <http://www.gbos.gm>).

## CHAPTER 6: DATA UTILISATION FOR GOVERNANCE IN AFRICA<sup>11</sup>

*Forms of government dramatically influence the community of practice in official statistics (Lehohla, 2008:4).*

Chapter six discusses the de facto situation with regard to data usage by African governments to assist their policymaking. Support for data management is premised on the notion that this empirical information is useful to governments, and this chapter elaborates on the extent of data usage for official planning on the continent.

### 6.1 The Role of Statistics Offices in National Planning

Empirical social research can support government accountability by providing the raw material for the development and monitoring of evidence-based policies. The collection of national statistics through survey research to inform government policy can be seen as an important component of governance. This is the primary reason for the existence of NSOs. The sharing of publicly funded government information, including survey information, is mandated in some western countries to support effective government. Data sharing is also seen to play a role in the economic growth of these countries.

International donor organisations foster the use of survey data for sound national decision-making and promote the sharing of public data in Africa for this purpose (Partnership in Statistics for Development in the 21st Century [PARIS21], 2009). Knowledge utilization is espoused as a governance principle in some countries. For example, the US government policy in this regard is that government information should be an accessible national resource. US Government administrators advocate cost-effective access to official information to promote sound governance (Johnson, 2005:1). Government agencies in the US are encouraged by national policies to share their information resources inter-departmentally and place this information in the public domain. They are not allowed to charge more than the cost of disseminating their information products and are encouraged to use the latest technology for speedy and efficient information distribution. However, it has been noted that

---

<sup>11</sup> An earlier version of this chapter was published in *Information Development* 25(1):22-32, 2009 entitled "Knowledge utilisation for governance in Africa: evidence-based decision-making and the role of SDAs in the region".

even in the US “administrators often make faith-based, rather than evidence-based decision-making” (Esanu and Uhler, 2004:70).

### **6.2 Evidence for Data Utilisation in Africa**

It is important, therefore, to investigate the assumption that, provided with relevant statistical information, governments in Africa will utilise this information for policy formulation. In order to advocate data sharing in Africa for effective policymaking in the region, it is vital to first confirm that social survey data is a useful tool in this regard. Does social survey research provide information relevant to the needs of African decision-makers, and, given the opportunity, do government policy-makers utilise empirical data for this purpose? It is possible that valid recommendations founded on sound social research are ignored due to a lack of will or understanding on the part of government functionaries. There is a need to investigate the obstacles to knowledge utilisation by African governments, including the attitudes of African policymakers to the use of research results to improve the quality of their decisions, and the level of skills among government planners in Africa to undertake this task.

Of interest too, are the factors that promote the utilisation of research results for government policymaking in Africa. It is postulated here that data sharing through SDAs could promote knowledge utilisation by African governments. However, data collection and data sharing for sound national planning needs a context where governments are receptive to the utilisation of research data, and have the capacity to utilise this data optimally. Before advocating for better quality data, and data that is more accessible, to inform policies in Africa there is a need to examine if governments, and specifically those in Africa, are making use of research for the formulation of policies. It is also necessary to investigate whether more and better data would encourage evidence-based decision-making on the part of governments in the region. Factors influencing the use of survey results for planning are relevant in this regard, and will be elaborated on in the next section.

### **6.3 The Gradual Influence of Research**

Studies of the utilisation of research in policymaking indicate that the impact of research on government policies is limited, and strongly curtailed by national political cultures and historical and institutional settings (de Gier, Henke & Vilgen, 2004:18-19). The realities around knowledge utilisation are that research findings do influence government policy, but indirectly

and gradually. One of the reasons for this is the complex nature of government planning. Policy-making is not a linear process of seeking information and utilising this information for problem solving. Rather, the many actors and interests involved result in a delay in the uptake of new information and novel methods. Thus research findings usually lead to incremental adaptations rather than immediate changes. The evidence gathered through research plays a role by gradually changing the viewpoints of policy actors and in this way influences subsequent policy decisions. Technical information and its analysis can be said to serve more of an enlightening than an instrumental function by providing a context for policy choice (Porter & Hicks, 2007:5).

Empirical research then, does have an influence on government policies, but its impact is mainly conceptual, and to a large extent “systemic” – that is, by shaping viewpoints gradually, leading to altered agendas which are reflected eventually in new policies. Unfortunately, the characteristic time lags between the publishing of survey results and their utilisation for policy formulation wastes research potential (de Gier, Henke & Vilgen, 2004:32-33). Formalised data sharing within and among African countries could expedite this process by making research data available in a more timely and accessible manner, and for comparative purposes, and this could result in the more regular and more efficient use of these countries’ limited academic resources.

### **6.4 The Influence of the Political Climate on Data Utilisation**

Empirical research is often used to support the existing policy agendas of governments (de Gier, Henke & Vilgen, 2004:32). Only when governments need solutions to intangible problems will they make use of research for conceptual input, and, in this context, Cross and others observe that input from academia can influence national policies (de Gier, Henke and Vilgen, 2004:31). However, decision-making will still be an organisational and political process, influenced by political events and vested interests (Porter & Hicks, 2007:3). Policy changes do not necessarily occur because new data becomes available. Available research may only be utilised after problem recognition by policy actors, hence the characteristic delay in the application of survey results for policy formulation. Policymaking combines expert knowledge with political influence, and this needs to be taken into account when examining the utilisation of empirical data by decision-makers in Africa.

An example supporting findings of the political influences on knowledge utilisation can be found in the history of South African survey research. During the Apartheid years state funded survey

research regularly collected demographic and income data, but only for its white population. The population census in South Africa, undertaken every ten years, only surveyed the white population. After the installation of the new democratic government, social survey research in the country burgeoned, partly as a result of demand from the new government for empirical data on all communities in South Africa for policy formulation (Seekings, 2002:1-3).

### **6.5 The Need for Consensus to Affect Data Utilisation**

Investigations have shown that research findings as possible solutions to social problems have to have gained a certain amount of consensus to proceed to the policy stage (Porter & Hicks, 2007:16). Principal Investigators on the KwaZulu-Natal Income Dynamics Study (KIDS), a survey project conducted in South Africa in 1998, found that, with regard to the influence of research on government attitudes, "the strongest impact is likely to occur when there is a measure of agreement between policy analysts" (John Maluccio, quoted in Bailey, 2005:537). Thus findings that are widely known and debated among academics and policy analysts have a greater chance of affecting policy changes. This could mean that the more widely research is distributed for secondary analysis the more consensus there will eventually be with regard to its value as a policy tool.

The dissemination strategy of the KIDS team included making the data available from SADA and the International Food Policy Research Institute (IFPRI) and via the World Bank's online African Databank (Bailey, 2005:530). The team promoted the findings through workshops and presentations to government departments and international development organisations. They found that placing their data in the public domain resulted in further analysis of the data, with the eventual consensus on its value resulting in the data having a greater impact on South African government policy (Bailey, 2005:536-7). SDAs like SADA can ensure empirical data is in the public domain and therefore available to provide solutions to problems when they are identified, and when the policy actors are ready to deal with them (Porter & Hicks, 2007:14-16). SDAs can assist in making research data available over the long term, and in a usable format, and ensure that it reaches a critical mass of commentators so it is thoroughly examined as a potential policy resource, allowing for the necessary consensus on its worth for government decision-making, and its eventual uptake into government policies.

## 6.6 Advocacy for Data Utilisation in Africa

The accessibility of research findings and publicity around these findings can be important factors influencing their future impact on government policy (de Gier, Henke & Vilgen,, 2004:32). However, even where research is available its utilisation for policy changes may be hampered by the established policy agendas of government practitioners. Studies suggest there is a need for intermediaries to ensure the effective functioning of the research-policy interface (Porter & Hicks, 2007:2). One level of intermediary exists in most cases of data utilisation, in the form of researchers who transform raw data into policy information for government practitioners. However, their policy analyses may need to be presented in an optimal manner to relevant government policy-makers to be used as a planning tool. Proactive dissemination of ideas to government decision-makers is necessary to affect policy-changes (Porter & Hicks, 2007:1-5). This type of advocacy plays an important role in facilitating the research-policy interface by ensuring research findings reach the attention of those government agents able to influence change.

A study undertaken by the AERC revealed the influential role individuals appear to play in facilitating research-policy linkages to ensure use of research findings in national planning (Hoffman, 1995:6-7). The investigators found numerous examples of effective research-policy interactions that existed only through the advocacy efforts of a single individual. They also noted breakdowns in these linkages through the departure of key actors in both the policy and research arenas (Hoffman, 1995: 7). An example of this is the active role key people in the Department of Science and Technology played in South Africa's participation in international initiatives related to the sharing of publicly funded research data (Page-Shipp, 2007). The value of advocacy work and the role of individuals in knowledge utilisation are also borne out in the history of the KIDS Study in South Africa. This research was utilised as a policy tool largely through the advocacy work of the principal investigators involved in the study. These researchers, by virtue of their positions on Government Commissions and Task Forces, were able to promote their research as a policy tool for the new democratic South African government (Bailey, 2005:530).

Advocacy work undertaken by researchers is unusual, however. Most academics believe their core competency lies in conducting the initial research, and presenting their findings in academic publications, which is necessary for academic recognition. The problem here is that academic writings are not created with government decision-makers in mind, and may thus not be read or

## DATA UTILISATION IN AFRICA

understood by them. A study conducted in 2004 in Cameroon revealed that Cameroonian policymakers in that country did not make substantial use of survey research even when this research was commissioned by them (Mbock, et al., 2004:41). One factor contributing to the non-utilisation of commissioned research was the mode of communication between researchers and government practitioners in Cameroon. Time constraints on policy-makers and other government practitioners meant they were unable to read the academic publications produced from research findings. These were written up by academics who were more interested in publishing their results in peer reviewed journals than convincing government agencies to adopt new policies. The information needs of policymakers were not taken into account with regard to the presentation of research results, and thus these went unused (Mbock, et al., 2004:40-41).

Information workers, such as librarians and data managers, may be able to play an advocacy role in knowledge utilisation on the African continent. Early attempts to create a dialogue between information practitioners and government policy-makers in the region did not meet with much success, however. An example of such an attempt is the work UNESCO's NATIS project undertook in sponsoring conferences and workshops in African countries during the 1970 and 1980s. This series of conferences and workshops was an attempt to bring librarians and government decision-makers together to raise awareness among the latter of the value of information in the policy-making process. Neill, commenting in 1991, highlighted their lack of success thus:

Government officials, planners and decision-makers exhibit an extremely low threshold of awareness with regard to the utility of information, and remain stubbornly unconvinced of its efficacy as a factor in the development process. The necessary conviction that would make NATIS work is not evidenced in the top echelons of government service with people who hold the purse strings (Nwalo, 2000:4-5).

However, advocates of evidence-based policymaking have begun to change the attitudes of African governments towards researchers, and decision-makers in the region have begun to appreciate the value of survey research for sound planning (Mbock, et al., 2004:38). Further advocacy is needed to ensure that the use of survey statistics becomes commonplace in government planning in Africa. Advocacy by intermediaries such as data archivists could result in a better understanding on the part of decision-makers of what social research can contribute to the governance process. The repackaging of information garnered by social surveys and the dissemination of this information through networked SDAs could ensure survey research provides

the necessary facts for policy analysts and advocates of policy change who have the skills to present this usable knowledge in a manner that indicates its policy relevance for governments.

### **6.7 Skills Levels in Government and Evidence-Based Policy-making**

The skills levels of policy actors within government could influence how social research is used in national planning. The research-policy interface may not be functional if government officials do not have the necessary skills to interpret and use social information. Government officials may also not understand the research process, which can lead them to make unreasonable demands on researchers (Bailey, 2005:539). Policymakers have limited opportunity and skills to read academic papers, and as noted above, researchers often do not aim their research reports at a policymaking audience. Lack of technical skills on the part of government planners can prevent the use of sound research for decision-making. Government functionaries generally receive national statistical information as reports or summary tables. The use of ICTs to disseminate relevant data could allow policymakers timely access to data to use as a planning tool. However, this would require a certain level of technical and analytical skill on the part of government functionaries.

A survey conducted in Ghana in 2003 investigated the perceived information needs and information usage of staff at District Assemblies (DAs) in the country through a series of interviews with DA staff members. Most respondents in the study had at least a first degree (94%), were computer literate (63%) and the majority (85%) used official statistics for planning purposes (Adams & Anum, 2005:140). Despite having relatively high levels of education, however, only 4% of respondents claimed to use the internet as a source of planning information. This low level of internet usage was linked to the fact that not all the District Assemblies in Ghana have internet access. However, it could also indicate a lack of the necessary competence in data analysis to make online data a viable option as an information source (Adams & Anum, 2005:141).

### **6.8 The Role of the Research-Policy Interface in Data Utilisation**

The type of research-policy interface in a country has been shown to influence knowledge utilisation by governments. The relationship between researchers as creators and interpreters of statistical data, and government decision-makers who utilise this data, has been shown to determine to a large extent how empirical information is used by national planners. A

relationship of mutual trust and understanding between government and academia seems to assist evidence-based policymaking, even if their agendas differ (de Gier, Henke & Vilgen., 2004:20).

### **6.8.1 Historical Animosity**

From the 1960s developments in Europe and North America were moving towards data sharing practices motivated by the mutual belief in the value of empirical research for sound governance and national development among policymakers and researchers (Scheuch, 1990:97). In Africa there has been slow uptake of research for decision-making and economic growth, and government support for empirical research to feed into policies has only recently been evident. Evidence-based policymaking in Africa has been hampered by a tradition of mistrust between government and academia, where researchers were at one stage viewed with suspicion for attempting to obtain information on governments whose operations were seen to be anything but transparent (Mwase, 1986:145).

The problematic research-policy interface in many African countries has been a barrier to knowledge utilisation for better governance in the region. In 1969 President Nyerere of Tanzania commented on this in a speech in which he expressed the hope that this would change:

We should abandon the idea that research is the same as spying, or that a researcher is really a person who is contributing nothing to our economy (United Republic of Tanzania second five year plan, 1969, quoted in Mwase, 1986:140).

In Africa, the un-co-operative attitude of government officials with regard to providing information for research meant that, in the past, even basic statistical data collected by government agencies, such as census information, was not easily available for independent analysis (Mwase, 1986:139). Academics were tolerated as long as they concerned themselves with theoretical research, that is, investigations unrelated to social conditions within their countries. In the 1970s and 1980s attempts to analyse social issues were met with government censure: for example, the 1970 closure of the Sociology and Philosophy Department of the University of Yaounde by the Cameroonian government in response to attempts by sociologists in this department to conduct research into social issues in Cameroon (Mbock, et al., 2004:38).

The 2004 Cameroonian study revealed that the majority of government-initiated research projects in that country did not result in any direct policy outcomes (Mbock, 2004:39). The

## DATA UTILISATION IN AFRICA

research showed that this poor utilisation of social studies for policy development resulted largely from the absence of any meaningful dialogue among academics, donor agencies and government practitioners. The study emphasised the need for communication among the agents involved in empirically-based policy creation, to counter the hierarchical nature of the government decision-making process and to assist academics in understanding the information needs of government agents:

In the absence of a space for dialogue and consultation, the policy-research interface operates in Cameroon in a climate of mutual indifference that is not devoid of suspicion (Mbock, et al., 2004:41).

Research by the AERC in 1993 and 1994 investigated the impact of survey research on policy-making in African countries. The AERC was established in 1988 to strengthen capacities in Africa with regard to policy research and training in economics, and the investigation was concerned with how economic research impacted on development policies in Africa (African Economic Research Consortium, 2009). The study examined the relationship between research and policy-making in four Anglophone African countries – Ghana, Tanzania, Uganda and Zambia, and three Francophone African countries – Cameroon, Côte D'Ivoire and Senegal. Information was gathered through interviews with senior government officials and academics in the selected countries, as well as employees in NGOs and the private sector. Opinions were also solicited from those working for international organisations – the African Development Bank, the two Communauté Financière Africaine Franc zone central banks, the International Monetary Fund and the World Bank. Interviews recorded respondents' experience of the research-policy interface in their countries (Hoffman, 1995:2). Although only economic policymaking was studied, the findings could have relevance for the role of empirical research in policymaking in Africa in general.

The study revealed that research had little direct impact on decisions taken by African governments. This was particularly evident in Francophone African countries, where historically university research was not expected to provide input for national planning. Government functionaries in Anglophone Africa were shown to have more interaction with university researchers, and thus more exposure to research results. However, with the exception of Uganda, university involvement in informing economic decision-making was shown to be minimal, thus limiting the practical application of university research. This research supports findings mentioned in section 7.3 regarding the slow uptake of research, which needs to be assimilated in a particular political milieu. In this case a further impediment to more research-policy linkages

was revealed to be the unstable situation at universities in the African countries investigated. This included financial difficulties and political unrest on campuses. These have a negative effect on both research output and the reputation of universities amongst government planners (Hoffman, 1995:6-9).

Views of researchers in the study were that government policy-makers were unenthusiastic about utilising the findings of local academics, and donor organisations tended to favour the work of expatriate researchers. Government functionaries held the view that allowing academics to suggest policy options could undermine their political authority, and resisted the possibility of critical commentary on decisions made. Policy-makers also felt that researchers did not understand the policy-making process, which sometimes involves decision-making in response to rapid political or economic changes. This ad hoc approach has been shown to lead to policy decisions being made which were contrary to those suggested by research findings (Hoffman, 1995:10-11, 14).

### 6.8.2 The Changing Research-Policy Interface in Africa

*If you can't measure it, you can't manage it.*<sup>12</sup>

The poor research-policy interface in many African countries in the past meant that valid empirical work carried out by academics went unused, wasting research potential and development opportunities (Mwase, 1986:144). However, in recent years, government attitudes in many countries in the region have changed from hostility towards investigations by academia to acceptance of the value of research input into government decision-making (Mbock, et.al., 2004:38).

An example from South Africa illustrates this. Critical examinations of social conditions during the apartheid years met with government censure. During the 1980s researchers at the Southern Africa Labour and Development Research Unit (SALDRU) at the University of Cape Town conducted an inquiry into poverty in the country, the results of which were published in 1984 in the report of the Second Carnegie Inquiry into Poverty and Development in Southern Africa (Southern Africa Labour and Development Research Unit, 2009). Widely covered in the local and international media, the findings from this research were vilified by apartheid-era policymakers. The then Prime Minister of the apartheid state viciously attacked the project and

---

<sup>12</sup> The then South African Finance Minister, Trevor Manuel, made use of this apothegm in his opening address to the inaugural African Symposium of Statistical Development., 2006.

## DATA UTILISATION IN AFRICA

its findings during debate in the South African Parliament (Wilson, 1996:227). In 1984 some universities were viewed by the Apartheid state as seditious institutions, and apartheid era governments did not value their research output. This led to the neglect of the key findings of the Carnegie researchers.

By contrast, survey results from a poverty study conducted by the same organisation at the University of Cape Town ten years after the second Carnegie study were welcomed by the newly established democratic government. The findings from this LSMS survey – the Project for Statistics on Living Standards and Development (PSLSD) - provided a guide for the new South African government's poverty-reduction policies from the first democratic elections in 1994. By this time relations in South Africa between academia and government had changed. Many of the influential practitioners in the new South African government had been involved in the liberation movement with those who later became academics. The subsequent improved levels of trust and communication between these researchers and the new government decision-makers could explain in part their willingness to translate this research into government policy.

As social surveys have become more complex, the role of the research-policy interface has become more significant in facilitating the utilisation of this type of information. Researchers have the skills to conduct analysis on this data to provide sound research findings for presentation to government planners. A relationship of trust is necessary to ensure the uptake of this research in government decision-making, as government functionaries seldom have the skills to conduct their own analyses, and need to rely on the abilities and opinions of academics when making policy decisions supported by empirical data. Thus research-policy interactions based on trust are vital for the proper implementation of evidence-based decision making in African countries.

In South Africa, beginning with the World Bank/SALDRU PSLSD survey of 1993-4, there has been a gradual increase in the use of empirical research by government agencies to inform national policy (Bailey, 2005:514-5). Subsequent survey projects in South Africa have successfully negotiated the research-policy interface to influence government action. This includes the 1998 KIDS survey which was a panel survey based on the KwaZulu-Natal households that were surveyed in the PSLSD survey.

The first wave of the KIDS survey was conducted in 1998, with subsequent waves in 2000 and 2004. While the first two waves were similar to the SALDRU survey in their exploratory nature,

the third wave of the study was expressly undertaken to provide input into government policy, that is “to improve evidence-based policy making for pro-poor policy through the analysis of [the] data” (Ben Roberts, KIDS team member, quoted in Bailey, 2005:523). Further assurance that the data would influence national policy was the decision by the South African Department of Social Development to provide funding for the project. The Department was interested in utilising the survey results to monitor the effect of their Child Support Grant, which had been introduced in 1998, coinciding with the first wave of the KIDS study (Bailey, 2005:524).

Concrete examples of the findings of the KIDS study being used to inform policymaking include its use by the Taylor Commission on Social Security Grants for the country, and in documents produced by the National Treasury, for example the 2001 National Budget Review (Bailey, 2005:536). Although usage of research results in government documents and national workshops does not necessarily translate into its direct input into policy changes, or the translation of these policies into government programmes, it does imply a willingness on the part of government to utilise research data for decision-making.

### **6.9 Data Quality and Evidence-based Decision-making in Africa: The Role of SDAs**

Political and cultural climates, and impediments to understanding between researchers as data producers and government decision-makers as potential data users, can restrict data utilisation by governments in Africa. This is exacerbated by other limitations on data use in Africa. Quality concerns such as limited data access may prevent data produced by African NSOs being used as a planning tool. For example the District Assembly staff participating in the 2003 Ghanaian survey complained that information provided by the Ghana Statistical Service (GSS) was not always easily accessible. 68% of respondents in this study claimed the statistical data supplied by the GSS was outdated. They were also dissatisfied with the lack of relevance of much of the data for their needs. For example, the GSS provided information at regional and not district level (Adams & Anum, 2005:141-143).

The establishment of national SDAs in African countries, where empirical data can be deposited for long-term preservation and access, can support data discovery for knowledge utilisation on the continent. SDAs could form a valuable component of the research-policy interface by facilitating access to research data for government functionaries and policy activists. This is already taking place in South Africa. For example DataFirst at the University of Cape Town

## DATA UTILISATION IN AFRICA

makes its data and services available on a subscription basis to staff from the regional Premier's office to assist local policymaking (Welch, 2007).

This chapter investigated the prevalence of evidence-based decision-making in African countries. Factors facilitating the use of survey information for government planning were examined, including the political climate and the role of data advocacy, skills levels in government, and the nature of the research policy interface in African countries. Government support for knowledge utilisation was shown to encourage the management and re-use of survey data for official planning. Lastly, SDAs were suggested as appropriate institutions to facilitate the use of survey microdata by African government planners. The following chapter details obstacles to the establishment of these and other data curating institutions for effective survey data curation in the region.

University of Cape Town

## DATA UTILISATION IN AFRICA



Photograph 9: The Liberia Institute of Statistics and Geo-Information Services (LISGIS). (Photo: LISGIS website, 2009: <http://www.lisgis.org>).



Photograph 10: The Publications Division of the National Statistical Office (NSO) of Malawi. (Photo: Malawi NSO website, 2009: <http://www.nso.malawi.net>).

## CHAPTER 7: OBSTACLES TO EFFECTIVE SURVEY DATA CURATION IN AFRICA<sup>13</sup>

*The present major task is ... to create ...infrastructural systems that are needed by the social sciences ... to utilise the vast amount of data ... that already exist. Today the social sciences ... are hampered by the fragmentation of the scientific information space. Data, information and knowledge are scattered in space and divided by language, cultural, economic, legal and institutional barriers (European Strategy Forum on Research Infrastructures (ESFRI), 2008).*

Obstacles to effective data management and data sharing in African countries have been identified in numerous investigations by international donor organisations involved in statistical capacity building in the region. These include the IMF's ROSCs available at this stage for eighteen African countries (International Monetary Fund [IMF], [2006?]). African NSOs have also provided information on constraints to data management and data dissemination as part of their NSDS (Partnership in Statistics for Development in the 21st Century [PARIS21], 2009). A survey of access to microdata from African NSOs was conducted by the author in March and April 2009 and the results were used to confirm and extend information found in these documents. A detailed report on the survey is included as Appendix D to provide further information on this aspect of data curation in the region. Survey findings are also listed, along with additional information on African NSOs, in the African Microdata Curation Index included as Appendix E.

### 7.1 Infrastructural Obstacles to Data Curation in Africa

One of the main impediments to effective data curation in Africa is the lack of political will among policy-makers in the region to use survey data for planning, and to encourage data re-use for economic development and policy monitoring. Government commitment would seem to be the principal requirement to foster this activity in the region. Given the scarcity of financial and skills resources in African countries, a pledge by governments is vital to initiate data curation processes and to ensure their continuity. Historically poor government support for this

---

<sup>13</sup> This chapter expands on issues raised in an article by the author entitled "A survey data archive network in Africa – possibilities and practicalities", published in *The IASSIST Quarterly* 31(1): 14-20, 2007.

activity means the regulatory, institutional and technological infrastructures for data curation do not exist in most African countries, and human resources have not been developed in this regard.

### **7.1.1 Problems with Regulatory Infrastructures**

Most African governments have yet to draw up comprehensive policies for the management and sharing of national data resources to support their development plans. Policies are necessary for formulating and initiating strategies for data preservation and dissemination. These should include the establishment of legal frameworks for the proper management and distribution of countries' data resources. This refers to official data, as well as data from other publicly funded research conducted in African countries. Legislation dealing with data confidentiality is also vital to ensure appropriate access to this resource. Currently policies related to data curation on the continent are few and far between. Most African governments do not have an official position on managing and sharing research data and, in fact, information management in general. This has led to a chaotic state of affairs in many African countries with regard to the production and distribution of information, including from official sources.

This situation is linked to resource shortages, but also to limited understanding among government decision-makers of the value of statistics as a practical tool for national social and economic development. Some African governments have in the past seen information as a propaganda tool, rather than a development resource (Yumba, 2002:239-241). Poor knowledge utilisation in the region results from African policymakers who are reluctant to use empirical data to develop policies and monitor their effects. To counter this state of affairs international donor organisations have encouraged the formalisation of data curation in Africa, at least with regard to official statistics. Donor support in this regard relates to their own data needs for programme monitoring, but is also based on their belief that data curation will advance good governance and economic growth in the region.

#### **7.1.1.1 Limited Focus on Data Curation in Existing Information Policies**

Existing policy initiatives of donors which incorporate data management plans include PRSPs, NICI plans and country NSDS. The PRSP process was initiated by the IMF in 1999 and PRSPs are comprehensive plans for poverty reduction required to support lending to IMF member countries. With regard to data curation, PRSPs include plans for statistical capacity building, including data management plans.

## OBSTACLES TO DATA CURATION IN AFRICA

The compilation of PRSPs extends the limited resources of African NSOs because these organisations are expected to provide comprehensive data for these reports. However, the creation of PRSPs helps NSOs to identify their assistance needs, and provides access to sources of funding to satisfy these needs (Achikbache, et al., 2002:160). The compilation of these strategy documents can also raise the profile of data curating organisations in African countries and encourage national commitment to data development. An examination of the PRSPs of African countries, however, shows that data curation is touched on only briefly in these reports, and the reports of some countries make no mention of statistical strategies (Chad, Malawi and Tanzania), although this is an aspect of infrastructural development emphasized by IMF and World Bank advisors in the process ([IMF], 2009b).

NICI policy documents formalise countries' plans for the utilisation of modern ICTs to achieve development objectives. These developed from initial attempts by the international community to support information management in developing countries. By 2007 thirty African countries had NICI plans in place. However, in most African countries these have not been integrated into other policy spheres such as technology or education, limiting their efficacy (United Nations Economic and Social Council, Economic Commission for Africa, 1999). These policies also need to be underpinned by government commitment to the development of key skills in the African workforce, and this is not always evident in the countries concerned (United Nations Economic and Social Council, Economic Commission for Africa, 2007).

Assistance to countries for the creation of their NSDS is another focus of donor support for information utilisation in the region. The creation of accurate information on statistical systems for inclusion in NSDS allows national and external assessment of countries' data capacities, providing a guide to their statistical capacity building requirements (Eele & Chinganya, 2005:25-26). By 2009 the governments of twenty six African countries were implementing NSDS and one country (Namibia) had data development as a component of its National Development Plan. 17 countries were working on new strategies for adoption or waiting for the adoption of newly created strategies.

That leaves just nine countries on the continent without a published formal statistical development strategy.<sup>14</sup> To some extent this progress reflects the keenness on the part of donor agencies to have countries "on board" this project, which suits their development agendas. However, the formation of formal data development plans indicates at least official

---

<sup>14</sup> Details of countries' NSDS are listed in the African data curation index (Appendix E) of this study.

acceptance of the value of data curation among countries in the region (Landais, 2008; Partnership in Statistics for Development in the 21st Century [PARIS21], 2008:1-3). Assisting the creation of data management policies is a key element in knowledge utilisation on the continent. An official document detailing strategies for data curation does not necessarily translate into effective data management for successful governance and sound economic policies. However, a formal policy is some indication of government buy-in to data curation principles.

### **7.1.1.2 Lack of Integration of Information Policies with Development Agendas**

Enabling policies for statistical development are vital to ensure the effective application of data resources in the region. The AAPA investigation revealed that by 2000 there was still limited sanction for data curation by African governments. Many African governments had failed to produce concrete national plans for statistical development, as recommended by the AAPA. The AAPA study noted that even where statistical development plans were formalised by African countries, these were not linked to national development priorities, and often failed to take local realities into account (Kiregyera, 2007:150).

The situation has improved somewhat since. In a number of countries these statistical development plans have been linked to national development priorities. For example, Mozambique has developed a NSDS which is linked to the country's National Poverty Reduction Strategy. Another example is that of Uganda, where statistical development plans centre on strengthening the Uganda Bureau of Statistics (UBOS) but are aimed at the development of the entire NSS (Simonpietri & Williams, 2005:13-14). These plans are also integrated with national growth strategies in other sectors in Burkina Faso, Kenya, Mali and Niger. However, in many African countries these initiatives are created at the level of country NSOs and in isolation from other data producers in NSS, and do not intersect with overall development plans Partnership in Statistics for Development in the 21st Century (PARIS21), 2008:12-13).

The policy-creation process serves to give prominence to statistical issues and international support, such as that provided by the PARIS21 advisory teams, can play a valuable advocacy role with African governments. However, comprehensive policies in this regard have yet to be put into effect in many countries of the region. Policies are often not all-encompassing but deal with specific components of data curation, such as official statistics or the technological

## OBSTACLES TO DATA CURATION IN AFRICA

dimensions of data management, and this limits their realization in these countries (United Nations Economic and Social Council, Economic Commission for Africa, 2007:iii-10; Partnership in Statistics for Development in the 21st Century [PARIS21], 2008).

### **7.1.1.3 Donor Driven Agendas**

Policy-making processes related to data curation are also limited by the fact that they are often donor driven rather than simply donor funded. Such interventions require a request from the governments of participating countries and have the support and cooperation of regional organisations such as UNECA, which represents African planning ministers. However, these programmes still often have a higher priority on the agendas of donor organisations than in the development plans of recipient governments and do not necessarily reflect genuine commitment to sound data management on the part of these governments.

Governments in Africa still do not regularly utilise survey data to improve the quality of their decision-making (Nwalo, 2000:5). This is changing, and African policymakers are beginning to see the value of the sound curation of national digital information to support social and economic development. For example, the Rwandan government is planning to establish a National Data Center to facilitate government institutions' access to and exchange of information (Rwandan National Data Center, 2009). Still, a data sharing culture does not exist in government structures in Africa, and few data management and data sharing policies have been set up in these countries. Active participation of government actors in data management plans is needed to change this. However, donor self-interest and lack of consultation between donors and government functionaries involved in data curation policymaking lead at times to a situation where local actors do not identify with policies sufficiently to ensure their efficacy or sustainability (Mbock, et al. 2004:44).

### **7.1.1.4 Minimal Private Sector Support**

Official sponsoring of data management is essential to prevent donor requirements becoming the dominant influence in data curation initiatives. A further problem with regard to funding of data curation in African countries is that the “triple helix” dynamic which is used in many countries to access funding and promote the technological developments needed to facilitate data sharing is unavailable in most African nations. The “triple helix” dynamic refers to the way

in which cooperation between governments, knowledge institutions and private companies can economically benefit countries (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2005:16). However, these partnerships cannot be readily set up in the region because in most African countries private sector institutions and universities do not have the necessary financial resources to assist governments with knowledge utilisation to aid national social and economic advancement.

### **7.1.1.5 Confidentiality Concerns in Policymaking**

Confidentiality considerations can also be used to prevent data sharing in the region. These are put forward by NSO directors as motivations for restricting access to national datasets, and weaken arguments for these institutions to place their microdata in the public domain. Government functionaries working with national statistics are accustomed to providing their data to a select group of policy-makers, donor bodies and researchers and are unfamiliar with a wider service orientation that includes the public in their client base (International Monetary Fund [IMF], 2004:3). In Africa, national legal frameworks need to be created to deal with issues of privacy and confidentiality, but specific government policy with regard to this aspect of data sharing has yet to be formulated in many countries of the region.

Middle managers in the majority of African NSOs are reluctant to take the initiative to provide national data to the public even where they see the value of data sharing as the typically bureaucratic management approach in these organisations prevents any initiative in this regard (Appendix D: African microdata access survey, 2009). With few exceptions, NSOs in Africa have no formal data access policies in place (Dupriez, 2008a). Decisions on access are ad hoc and it is sometimes informal contacts that will enable access to the data. Requests for data are either declined or handled on a case-by-case basis, with this approach being justified by confidentiality concerns (Appendix D: African microdata access survey, 2009). The former stance wastes national resources by preventing reuse of data by researchers for policy analysis to aid better governance, or to provide innovative input for economic growth. The latter position can lead to onerous requirements for data access, again restricting usage and limiting the benefits of knowledge utilisation for African countries. The absence of clear policy guidelines on data protection in African countries means government functionaries will continue to resist making decisions in this regard, for fear of overstepping their authority.

### **7.1.1.6 Other Hindrances to Effective Data Curation Policies**

Even where African governments have shown commitment to data sharing practices, other constraints hamper the implementation of data access policies. Political instability in some African countries leads to policy re-direction with each change of government, and therefore there is no time for any information policies to become practice in these countries. The lack of financial resources and expertise to implement policy proposals is a further restricting factor. In some countries in Africa the result of these impediments is that legislation has not kept pace with technological developments. For example, the legal deposit laws in many African countries only deal with printed publications and do not make provision for digital information sources (Yumba, 2002:241). In many African countries legislation and government programmes aimed at improving access to digital information are hampered by lack of enforcement mechanisms or outright exemption of government publications (Shibanda, 2006:7). A further difficulty is that the implementation of data curation policies in African countries relies heavily on the work of local data advocates, rather than being the product of a national data sharing culture in these countries. For example, in South Africa the process of developing policies related to the sharing of research data is being held up by the loss of key government personnel (Page-Shipp, 2008).

### **7.1.2 Problems with Institutional Infrastructures**

Unlike many countries in Europe, data curation in African countries takes place almost exclusively at the official level. Any study of data curation constraints in Africa will therefore need to focus mainly on impediments to the management and sharing of data by African NSOs.

#### **7.1.2.1 African National Statistics Offices**

In Africa NSOs are the key data curation agencies within country NSS and are responsible for coordination of data collection and data management within these systems. NSOs form part of the country's civil service, either as autonomous government departments or as departments under a body such as the Ministry of Planning. NSOs are mandated to collect economic, demographic and social statistics to support government planning (Ching'anda & Ntozi, 1998:235). However, there is no culture of data archiving or data sharing at government institutions in Africa (Committee on the Geographic Foundation for Agency 21, Committee on Geography, Mapping Science Committee and the National Research Council, 2002:5). Even

## OBSTACLES TO DATA CURATION IN AFRICA

in specialist preservation agencies such as state archives, the emphasis is on long-term storage of information sources, rather than on curation for intended future use.

A lack of government interest in the merits of data curation leads to few resources being made available to NSOs for data sharing (Kiregyera, 2005:71). Despite government lip-service to the value of evidence-based policymaking, and official claims regarding commitment to harnessing empirical data for economic growth, scant government funding is allocated to NSOs in African countries. NSOs in most countries of the region are chronically underfunded and suffer from shortages of basic equipment such as computers and vehicles (Manning, 2006:1-2). Skills shortages and high staff turnover due to low salaries also result in a paucity of analytical expertise in these institutions (Lufumpa & Mouyelo-Katoula, 2005:31-32). Government expenditure on statistics in African countries is mainly allocated for data collection, and very little funding is made available to NSOs in the region to support the long-term preservation and sharing of national data (Kiregyera, 2005: 70-72; Moultrie, 2008). The outcome of this is that in many NSOs data archiving is not practiced in a systematic manner, which has led at times to data losses or the production of unreliable data (International Monetary Fund [IMF], 2004).

Even where NSOs have participated in donor projects to advance the quality of their data, statistical output has been compromised as a result of these resource shortages. A good example of this is the variable results from African participants in the IMF's GDDS project. Participation in the GDDS Project is an indication that the country NSS subscribes to a minimum set of standards with regard to the production, preservation and dissemination of their official data. National coordinators appointed by African governments to work with the IMF have generally been senior-level managers from NSOs or central banks, indicating that participating African countries take the initiative seriously. Still, the work of the IMF and country coordinators has not been able to overcome the dissemination problems of much of the data produced by African NSS. In Africa adoption of production and dissemination best practice with regard to survey data has been slow. The data quality initiative originally envisaged that countries would progress from the GDDS to the SDDS once adherence to the GDDS standard had been achieved for several years. However, only three countries in Africa have made this progression – Egypt, Morocco and South Africa (all in 2005) (Alexander, et al., 2008:61).

Even where international funding provides assistance this can be problematic. In some African countries 70 percent to 80 percent of NSO budgets are donor-supplied and this imbalance in

## OBSTACLES TO DATA CURATION IN AFRICA

funding sources can result in survey projects being donor rather than country-driven, again compromising data relevance. Data timeliness is also affected by resource shortages in official data curation institutions. In many African countries funding constraints mean surveys are conducted irregularly and periods between population censuses are unacceptably long. There is thus a paucity of recent survey microdata for repurposing. Preservation issues plague these institutions as they make do with outdated hardware and software. Often historical survey data is lost through lack of resources to migrate the data from storage on legacy equipment (Moultrie, 2008).

The poor provision of resources to African NSOs impedes data discovery on the continent. African NSS are decentralised with data being produced by a number of organisations, including official statistics agencies, other government departments, academic institutions and donor organisations. No systematic inventory of data available across institutions in NSS is undertaken. The data co-ordination role mandated for NSOs in African countries is seldom accomplished, as these agencies have barely enough human and financial resources for their primary task of producing national statistics (Kiregyera, 2005: 71-72). NSOs also have limited funding and technological resources to establish effective user-producer communication channels. This leads to inadequate links between these data producers and data users in African NSS. These systems often end up being supply rather than demand driven and this affects the relevance of the data produced.

This problem was confirmed by a survey conducted by the author in order to investigate the availability of survey microdata from African NSOs for repurposing. An obvious indicator of national data producers' commitment to providing accessible, user-friendly and relevant statistics is the extent and effectiveness of their public communication channels related to data dissemination. The microdata access survey was therefore undertaken using contact points on the websites of the NSOs. A request for was emailed to each NSO and where there was no response the data request was re-submitted by fax.

The following obstacles were encountered: NSOs in eight of these countries did not have websites or had websites which could not be accessed (Angola, Comoros, Democratic Republic of Congo, Eritrea, Libya, Mali, Somalia and Zimbabwe) Twenty-two of the countries could not be contacted, with both email and fax contact points not functioning (Algeria, Benin, Burundi, Burkina Faso Central African Republic, Chad, Congo, Djibouti, Ethiopia, Gabon, Gambia, Guinea, Guinea-Bissau, Liberia, Rwanda, Seychelles, Sierra Leone, Sudan, Swaziland, Togo and

Zambia). For thirteen countries email and fax details were available and both used to send data requests, but these elicited no replies (Cameroon, Cape Verde, Côte d'Ivoire, Egypt, Equatorial Guinea, Lesotho, Madagascar, Mauritania, Morocco, Niger, Tanzania and Tunisia). Procedural and technical impediments hampered follow-up by NSOs in three countries (Malawi, Namibia and Uganda).

In two countries NSOs were contacted, but staff of these NSOs provided the information that they did not share their data for research (Mozambique, São Tomé). These institutions currently have no microdata access policies that enable them to supply microdata files, although in the case of Mozambique they have the technology available to place the data in the public domain (a NADA catalogue).<sup>15</sup> NSOs in nine African countries had functioning communication channels in place that led to the opportunity to obtain microdata files. These included Mauritius, where the NSO provides data for research purposes, but under onerous conditions. The Mauritian Central Statistics Office required payment of 4000 Rupees (\$129/R988) and a visit to the country to swear an oath of secrecy before the Director of Statistics, or the appointment of a representative to undertake this on the researcher's behalf. The Ghana Statistical Service required payment to obtain microdata files (\$320/R2 446 for the microdata from a panel study in three waves, from 1997 to 2006). This is not costly for the amount of data provided, but definitely represents an obstacle to access for researchers from smaller, less well-funded research institutions on the continent. The National Agency of Statistics and Demography in Senegal responded to a request submitted via a NADA online form with an invoice for 411 331 FCFAs (\$902/R6 897) for the data from their 2005 poverty survey, which represents a substantial sum for any African researcher.

Finally, four NSOs provided relatively obstacle free access to the microdata from their national surveys (Botswana, Kenya, Nigeria and South Africa). In response to an emailed and a faxed request, the Botswana Central Statistics Office (CSO) and the Kenya National Bureau of Statistics (KNBS) emailed the requested data files and supporting documentation. The Nigerian Bureau of Statistics (NBS) and Statistics South Africa (StatsSA) allow downloading of their microdata files via their websites, the NBS through their NADA web-based catalogue and StatsSA via their online NESSTAR server (African microdata access survey 2009: Appendix D).

---









<sup>15</sup> A subsequent meeting with the President of the Mozambican INE, João Loureiro, revealed that the NSO provides their microdata to the Mozambican Department of Science and Technology to make available to researchers. However, this was not the information initially supplied during the survey. This information will be followed up for a future publication based on this survey.

## OBSTACLES TO DATA CURATION IN AFRICA

**Map 1: African Microdata Access Survey 2009**



The Circle indicates the NSO has the IHSN's NADA web-based microdata management software installed. The dot indicates the NSO has AASDA membership.

Category 1	No website access	
Category 2	No contact	
Category 3	No reply	
Category 4	No policy	
Category 5	No follow-up	
Category 6	Onerous requirements	
Category 7	Data available on payment	
Category 8	Good access	

### 7.1.2.2 Absence of Other Data Curating Institutions

The paucity of data curation institutions in African countries reflects the poor research infrastructures and weak research-policy interfaces in these countries generally. Historically few structures to support empirical research and scientific analysis have been developed in the region. Empirical research conducted in colonial times was mainly concerned with the collection of data to support the functions of the colonial administrations and no attempts were made in that period to build local statistical capacities to assist development needs. Post-colonial governments oversaw the establishment of national higher education systems and research efforts were geared towards national development. In the 1970s and 1980s there was a boom in the development of national scientific and educational institutions in the region. During the 1990s, however, there was a decline in the overall research capacity of African countries. The current situation reflects this decline, with a lack of support and institutional direction for research from national coordinating bodies in the region (United Nations Educational, Scientific and cultural Organization [UNESCO], 2005:178-9).

The consequences are low levels of research output, and few skills in appropriate data analysis (United Nations Economic and Social Council, 2001:13). The systems in place in other regions to share empirical data from social surveys, including networks of SDAs, do not currently exist in Africa. South Africa is the exception, as the South African government has established SADA to facilitate the sharing of South African survey microdata (Lesaoana, 1997:6). There are no similar research support institutions in other African countries. Consequently much data from surveys conducted by national and international organisations in Africa is not available to analysts apart from initial research teams.

Statistics are used by several groups in African countries, including government ministries, research and training institutions, the media, civil society groups and regional and international organisations. The limited opportunities for interaction between these data users and national data producers in African countries prevent feedback for the improvement of data quality. This is an impediment to the effective repurposing of African data. The shortage of institutions in the region that link data producers to the users of the information they generate is detrimental to scientific development and effective policymaking in the region (Marshalling technology for development, 1995:30). Often African data resources are collected locally and archived

outside the continent and the microdata may be difficult to access for researchers in African institutions.

### 7.1.3 Poor Technological Infrastructures

Enabling government policies and a network of well-resourced institutions on the continent require underlying technological infrastructures to affect data curation. Many countries in Africa lack the basic telecommunications infrastructure for providing research support of this nature. Technological restrictions on data curation are infrastructural problems, as they refer to inadequate ICT infrastructures in African countries for data sharing. These are also institutional constraints on data repurposing, as data curation institutions are dependent on technology to function effectively. Policy issues are also relevant here, as policies are required to establish enabling technological infrastructures for proper data curation in countries of the region. Other policy dynamics related to technology include governmental control of telecommunications in African countries (Dutta & Mia, 2009:47,121). Thus the necessary conditions for the building of Africa's technological readiness to make effective use of national data are dependent on overcoming other obstacles discussed in this study.

Poor technological infrastructures on the continent include unreliable power supplies in most countries. National electrification levels are generally low in Africa, with the exception of North African countries and Southern Africa. A study conducted by UNECA in 2007 to assess the power sector in selected African countries showed that in the majority of the fourteen countries surveyed, even in urban areas fewer than half the households had access to electricity. System losses in these countries were also high, up to 41%. This did not compare favourably with the international target of losses of 10-12% (United Nations Economic and Social Council, Economic Commission for Africa and United Nations Environment Programme, 2007:53-59)

Data curation has to be accomplished through the medium of modern information and communications technology. ICTs are necessary for data sourcing, for example through regional web-based catalogues, for data preservation, for example on file servers, and for data dissemination, for example via the internet within a regional data sharing network. The proper utilisation of ICTs, particularly the internet, is the foundation for data curation in the region (Kirkman, et al., 2002: 254, 310, 280). However, African governments are not currently taking advantage of the data management opportunities afforded by modern information and

## OBSTACLES TO DATA CURATION IN AFRICA

communication technologies to set up ICT infrastructures in their countries despite this being a prerequisite for any continental data sharing system. Impediments to the establishment of enabling technological infrastructures for data archiving and data sharing in Africa include policy limitations, and issues related to hardware and software use and access to the internet.

There have been recent attempts to develop the Information and communications infrastructures of African countries, but Africa is still the continent with the most poorly developed ICT capacity (New Partnership for Africa's Development, 2002:3). South Africa has an advantage over its neighbours with regard to the establishment of data sharing networks, in that it has a fairly well developed technological and research infrastructure. An indication of this is its comparative position on the Networked Readiness Index (NRI). The NRI is a scale developed at Harvard University which attempts to measure countries' current and potential ability to utilise modern ICTs for their advancement. Both the level of development of a country's networks and factors enabling future ICT growth are examined in the compilation of this index. Factors taken into account in the NRI are those which are judged to promote network development and network usage. These include the availability of hardware and software and the current percentage of internet users in a country (internet penetration rate) (Dutta & Mia, 2009:3-8).

African countries have poor ratings on this index of networking capabilities. Only Tunisia (38th) is ranked in the top fifty countries and only eleven out of the thirty one African countries listed make it into the ranks of the top one hundred. These are Tunisia (38), Mauritius (51), South Africa (52), Egypt (76), Botswana (77), Senegal (80), Morocco (86), Nigeria (90) The Gambia (91), Namibia (92) and Kenya (97) (Dutta & Mia, 2009:3-10). Findings of the survey on access to government microdata conducted in 2009 for this study indicate that these low rankings are justified, as technological and skills hurdles seem to be at the heart of inefficient data exchange in the region (see Appendix D).

A problem linked to poor technological infrastructures in the region is the low internet penetration in African countries. Efficient data sharing in the region relies on the internet. The establishment of networks for survey data sharing is not possible without internet access (Nwalo, 2000:6). Internet connectivity on the continent is growing – currently all 53 African countries have some form of internet linkage (Internet Systems Consortium, 2009). However, internet penetration in Africa is still very low. Even in South Africa, which has a reasonably strong research infrastructure and ICT resources available to the general public, internet usage is low by international standards. A 2008 survey conducted in selected African countries revealed low rates of access to both computers and the internet in the African countries examined (Gillwald &

## OBSTACLES TO DATA CURATION IN AFRICA

Stork, 2008:25-27). Survey findings indicate computer ownership ranging from 0.22% in Ethiopia to 5.47% in Kenya. Figures for internet use, within and outside the home environment, range from 0.7% in Ethiopia to 15% in Kenya and South Africa. The average broadband penetration rate in OECD countries is 18.8% (Southwood, 2008:1).

Poor exchange rates due to weak currencies in African countries mean that prices of hardware and software can be prohibitive in many countries in the region (Nwalo, 2000:5). African institutions, including government agencies, often have a shortage of funds for computer and communication infrastructure leading to a situation where computer facilities are shared within organisations, limiting access to the internet. Low internet penetration in Africa is also partly a product of generally low levels of literacy in countries of the region, including computer literacy. However, the impressive uptake of mobile phone technology in Africa indicates that affordability and usability are the key influences on internet penetration in this context (Dutta & Mia, 2009:28).

The high cost of internet access in Africa can thus be seen as a key factor inhibiting the establishment of a critical mass of competent internet users to develop data sharing capacities (Mutula, 2003:489). The average retail price for basic broadband in sub-Saharan Africa in 2006 was US\$366 a month, while in India it was US\$44, in Europe US\$40 and in the United States US\$12 (Williams, 2006:i). The cost issue needs to be addressed by governments prepared to fund access as personal investment in internet access is not an option on a continent where many people cannot afford even a phone line (Nassimbeni, 1998:160).

The submarine fibre-optic cables currently under construction or recently in operation on both the East and West coasts of Africa will undoubtedly lower the cost of internet access in African countries (*African undersea cables*, 2009; Southwood, 2009:1; Williams, 2006:4). An example is the Seacom owned cable connecting Kenya, Mozambique, South Africa, Tanzania and Uganda to Europe and Asia, which went live in July 2009 (*East Africa gets high-speed web*, 2009:1). However, prohibitive internet costs in Africa are exacerbated by the telecommunications regulatory frameworks in many African countries. Out of Africa's fifty three countries, over half (29) still have incumbent telecommunications companies that are majority government-owned (*Financial tide goes out on incumbent telco privatizations*, 2008:1). The telecommunications infrastructure and how it is managed directly affects networking opportunities in a country. Government monopolies limit infrastructure development in African countries, and prevent access

## OBSTACLES TO DATA CURATION IN AFRICA

by competitors to what could be cheap technologies, if utilised on a larger scale (Beaven & Martin, 2004:16).

State-owned telecommunications companies also often lack the financial resources for provision of new services. They may also react too slowly to the rapid technological advances in the sector United Nations Economic and Social Council, Economic Commission for Africa, 1999; Mutula, 2003:492). There has recently been an easing of monopolies in African countries' telecommunications sectors. An example is the introduction in South Africa of a second telecommunications operator, Neotel, in 2006. Previously this sector was controlled by the part state-owned company, Telkom (Neotel, 2009). The introduction of competitors in this sector has been linked to increased internet penetration in African countries (Kirkman, et al., 2002:121).

Overcoming technical barriers to data archiving and dissemination is a necessary condition for the effective implementation of data advancement programmes in Africa. The software tools designed by the IHSN and provided to African NSOs to enable data curation best practices at these institutions have been a valuable resource in this regard. Still, the poor technological infrastructures of many of these countries hamper the implementation of these resources and prevent their optimal application for data sharing in countries of the region (African Microdata Access Survey 2009, Appendix D).

### **7.2 Human Resource Constraints to Data Curation on the Continent**

Human resource constraints to effective data curation relate to a shortage of IT professionals, survey statisticians, and practitioners trained in survey data management. There is also a paucity of African researchers competent in secondary analysis of survey data.

#### **7.2.1 Shortage of Skills to Support ICT Infrastructures for Data Sharing**

Many African countries lack the educated and skilled workforce needed to take advantage of the new information and communications technologies to utilise survey data as a national resource. This includes people with management and technical skills to administer any technological infrastructures (Kirkman, et al., 2002:281). African countries have few computer engineers and technicians (Nwalo, 2000:6). These are needed to ensure the proper functioning of the technology to enable data curation. Skills shortages in this area result in

## OBSTACLES TO DATA CURATION IN AFRICA

problematic communication between data producers and potential users of the data, as well as dissemination problems. For example at the National Institute for Statistics, Economics and Demographic Studies (INSEED) in Chad there is only a small budget for the website, with no financial resources allocated to training for website management. This lack of skills and lack of opportunity for training affects data discovery at this institution. The INSEED website does not support data discovery and the metadata for surveys conducted by this NSO can only be located on the GDDS website (International Monetary Fund, 2009).

### **7.2.2 Inadequate Statistical Training**

Statistical training in Africa takes place formally at universities or regional STCs. These institutions also conduct short courses and workshops to advance statistical skills in the region. However, university-level training for statisticians is a low priority in African countries. Regional STCs were therefore established by donor organisations to fulfil the need for trained statisticians in Africa. However, STCs in Africa usually do not include training in official statistics in their curricula (Lufumpa & Mouyelo-Katoula, 2005:31). This applies also to data management training for statisticians as courses offered at STCs do not take into account the need for archiving strategies for official statistics to ensure their long-term preservation and distribution.

STCs, as key training institutions in the region, receive international funding. For example, from 2003 to 2005 EASTC, ENEA and ENSEA were funded by the IMF to the amount of US\$2.2 million. However, international financial support for Statistical Training institutions in Africa has declined in recent years, and funding for scholarships for African students to gain skills abroad has all but ceased. (Partnership in Statistics for Development in the 21st Century (PARIS21), 2006:7). This further exacerbates a situation where there is an acute shortage of human resources to advance statistics in the region.

The ADB IMF and World Bank statistical capacity building programmes conduct training workshops but these are sporadic. Non-governmental training institutions exist in some African countries, as do governmental training institutes outside NSOs which could assist in training, for example in Algeria and Egypt. However African NSOs often do not have adequate plans or budgets to send staff on external training courses (MEDSTAT II. 2008:25-26). Also problematic are the daunting bureaucratic procedures required at some NSOs to enable staff to attend external training courses (International Monetary Fund [IMF], 2002:82).

Consequently most statistical training of NSO staff is undertaken in-house. This training is conducted on an ad hoc basis and limited resources and work overloads mean a lack of enthusiasm for these by staff of NSOs (Bernardo, 2008). There are generally no formal systems in place at African NSOs for collecting information on training needs to feed into future training programmes. Planning and coordinating between NSOs and relevant organisations, such as statistics departments of national universities, is rare in African countries. NSOs also do not coordinate with other government statistical producers to assess the overall training needs of the NSS (United Nations Economic and Social Council, Economic Commission for Africa, 2001:16).

African data curation is also poorly represented by professional associations. Currently there is little interaction among African statisticians in the region. The African Statistical Association no longer exists as funding shortages led to this body disbanding in 2000. A study commission by the ACBF in 2008 recommended the re-establishment of the African Statistical Association because a strong network of practitioners was perceived as necessary to advance development in this field (African Capacity Building Foundation, 2008:2).

### **7.2.3 A Paucity of Data Management Skills**

Survey data management is such a novel task in Africa, that there are only a handful of practitioners in the field. Although surveys have been conducted since colonial times, the preservation of survey results as microdata files has only been made possible in the last decades, through advances in technology which have facilitated the efficient and cost-effective storage and sharing of this information. Decision-makers in African countries have only recently begun to see the value of long term preservation of data, and its management and dissemination according to international standards to facilitate its repurposing for research and policymaking. They have thus not supported the introduction of any national training programmes to foster skills in this regard. Established University courses in data management are part of the curriculum in some European and North American universities, for example the course in Social Science data librarianship at Cornell University in the US (Olken & Gey, 2006). Similar courses do not exist in Information Science or Statistics Departments of African universities and colleges (Sele matsela, 2009). Data management training in Africa takes the form of short courses or workshops, or technical assistance as part of international statistical advancement programmes such as the IHSN/ADP project.

### **7.2.4 Limited Research Capabilities**

Effective data curation in Africa depends on the existence of a critical mass of researchers to effect data repurposing in the region. The low density of scientific activity in Africa is a limiting factor in data use. This results from a lack of government support for research in many African countries, due to budget constraints but also to the low regard for learning held by government policy-makers in some African countries (Nwalo, 2000:6). For example, in 2002 Africa's gross expenditure on research and development (GERD) represented only 0.4% of world expenditure and South Africa was responsible for 90% of this expenditure (United Nations Educational, Scientific and Cultural Organization (UNESCO), 2002:187). A resultant problem is the "brain drain" of researchers from African countries who find more rewarding employment overseas (Kirkman, et al., 2002:196,254,280). This situation is due to low salaries, as well as insufficient resources for training and education in African countries (Bits of power, 1997:41-42; Lufumpa & Mouyelo-Katoula, 2005:31-32).

A shortage of data analysis skills among potential data users, including those in government and private companies, means the absence of researchers to lobby for the use and availability of data resources. Committee on the Geographic Foundation for Agency 21, Committee on Geography, Mapping Science Committee and the National Research Council, 2002:5). African universities do not offer courses in data management and, except in South Africa, university courses in data analysis are not aimed at future policy-analysts and policymakers (Bailey, 2005:540). The paucity of data discovery resources and data sharing possibilities in African countries mitigates the use of survey microdata in social science curriculums at African universities. This is exacerbated by and contributes to the low levels of empirical research in the region.

### **7.3 Poor Co-ordination among Stakeholders in Data Exchange in Africa**

Role-players in African data curation include government agencies and policy-making and regulatory bodies, intergovernmental scientific organisations, publicly funded data management institutions such as data centres and libraries and publicly-funded research institutions such as universities (Committee on Issues in the Transborder Flow of Scientific Data and National Research Council, 1997:63). Private sector institutions may also be data producers and data users, and need to be taken into account in any data curation strategies established at international or national levels.

### **7.3.1 Problems with International and Regional Coordination**

The United Nations Statistics Division (UNSD) is the umbrella body for country NSS. Funding for UNSD projects comes from the UN and development partners such as the PARIS21 Consortium and the World Bank. Regional statistical supervision is carried out by UNECA and its ACS (United Nations Economic and Social Council, 1997). Donor agencies have made resources available for data development projects in African countries. For example in 2004-2005 alone 10.2 million dollars was spent on statistical capacity building projects in Sub-Saharan Africa. However, these data advancement initiatives have in the past been introduced in a piecemeal fashion, and have lacked sustainability due to limited government support and the changing priorities of donor organisations. These problems have led to duplication of effort and disappointing project results (Beaven & Martin, 2004:14; Partnership in Statistics for Development in the 21st Century [PARIS21], 2006:22-23).

### **7.3.2 Problems with National Coordination**

At a national level statistical management in Africa is the responsibility of NSOs, which are also legally mandated to coordinate country NSSs. These systems are currently characterized by decentralized statistical production and poor communication among data producers, and little dialogue between data producers and data users (African microdata access survey 2009; Bernardo, 2008). Funding for NSS comes from government sources, as well as UN agencies and their development partners. Funding from governments is inadequate, however, with 70 percent to 80 percent of NSO budgets in some African countries funded by donor organisations (Lufumpa & Mouyelo-Katoula, 2005:31-32). NSOs lack the financial and staff capacities to affect their co-ordination role within the NSS. The plans of African governments for national statistical advancement need to take place within the context of global statistical agendas in order to be effective, but this has not always been the case.

Statistical development in Africa also suffers from poor user-producer coordination. African NSOs are legally mandated to establish user/producer forums to support coordination of statistical development within countries (Owino, 2007:11). This is one of the key recommendations of the Addis Ababa Plan of Action. However, this was only implemented in some African countries, and some of these have become dormant (United Nations Economic and Social Council, Economic Commission for Africa, 2001; Kiregyera, 2005: 71; Owino, 2007:19-20). A survey conducted for this research found that NSO user-producer communication channels in African countries were not effective. For example, the

## OBSTACLES TO DATA CURATION IN AFRICA

contact details provided on NSO websites were outdated for many of these organisations. The majority of these organisations had no policies and procedures in place for dealing with user requests (African microdata access survey 2009, Appendix D).

Inadequate co-ordination at the national level includes the continuing poor research policy interface in many African countries. For example, a report published in 2000 by the Centre for Development and Enterprise – a think-tank representing large companies in South Africa - found that there was no formal communication mechanism for the exchange of ideas between academia, industry and government in South Africa. Such structures are utilised in some countries to promote dialogue among these groups to support research and development. For example the Council for Industry and Higher Education plays this role in the United Kingdom (James, 2001:91).

This chapter dealt with obstacles to the effective management and sharing of survey microdata in African countries, including the paucity of regulatory, technological and institutional infrastructures to affect data curation in the region, which is linked to the low regard still shown by many African governments for the use of empirical research as a planning tool. Skills shortages are also shown to hamper data usage in African countries and the poor research and training infrastructures in these countries do not provide the necessary competencies to support data curation. The concluding chapter will summarise the findings of this study and make recommendations for the creation of an enabling environment for effective data curation in the region.



Photograph 11: The Lesotho Bureau of Statistics (LBOS). (Photo: Lynn Woolfrey, 2008).



Photograph 12: NADA Training at the Lesotho Bureau of Statistics (LBOS). (Photo: Lynn Woolfrey, 2008).

## CHAPTER 8: CONCLUSION AND RECOMMENDATIONS

*The political will of governments is vital for any decision-making regarding information management (James, 2001:156).*

As this research has demonstrated, the efficient management and sharing of national survey data can provide governments with valuable information resources to support their development planning. In attempting to answer questions around effective data curation strategies for African countries, this study has shown that, worldwide, attempts to manage survey data as a national resource have faced a number of issues. The historical overview detailed in chapter two reveals that international efforts in this regard have been confronted with funding dilemmas, as well as ethical concerns such as those around ownership and control of data and data confidentiality. Data quality issues have also influenced the development of data management institutions and practices worldwide, as elaborated on in chapter four of this study.

### 8.1 Conclusion: Political Facilitation for Data Curation in Africa

A review of obstacles to survey data curation in Africa which forms chapter seven of this thesis shows that similar issues arise with regard to data curation in African countries. Funding constraints on data archiving and confidentiality restrictions also preoccupy African data managers. However, restrictions on data usage are compounded in African countries by poor technological infrastructures and skills shortages. This chapter and chapter six which deals with official data utilisation in Africa also confirm that a major impediment to effective data curation in the region is the lack of government appreciation for evidence-based decision-making which leads to meagre resources being allocated by African governments for the management of data for reuse.

The historical account of survey data curation in Africa detailed in chapter five investigates strategies adopted by international donor organisations working with African governments to overcome these barriers to survey data usage. These have included the building of country regulatory, technological and institutional infrastructures to support data curation, and training initiatives to increase data curation capacity on the continent. As an investigation of appropriate mechanisms to affect data curation in African countries for data repurposing this study has revealed that the above strategies have had limited success in promoting this activity in the region.

## CONCLUSION AND RECOMMENDATIONS

Chapter seven of this study shows that the statistical capacity building projects of donor organisations will have limited success without the active involvement of African governments. Using evidence provided from international data management best practice, as well as the results of a survey on microdata access in African countries (Appendix D), this study has demonstrated that a key requirement of effective data curation is governmental support. Optimising the financial and technical assistance of the international community and national data utilisation potential requires leadership rather than simply governance by African policy-makers who appreciate the value of information as a development resource and are committed to its usage for the betterment of their societies.

### **8.2 Recommendations Regarding Governmental Support for Data Curation**

The investigation of data sharing practices in African NSOs conducted as part of this thesis (African microdata access survey: Appendix D) shows that the absence of this government support, and the concomitant lack of data curation policies, hampers data usage in the region. The comparative component of this study demonstrates that where there is appreciation on the part of government decision-makers for information as a national resource this will be translated into the creation of data harvesting and data sharing policies. Data curation policies of governments in Africa could be formally articulated as the digital information management component of National Information Policies, where these exist, and could formalise and institutionalise the changes necessary to effect data curation in these countries.

#### **8.2.1 Policies for Data Curation Advocacy**

Data curation policies could initiate strategies to encourage the development of a culture of data sharing in countries of the region. An examination of the role of data advocacy in data utilisation undertaken in chapter seven provides evidence for the importance of data advocacy for changing mind-sets among government functionaries and the general public with regard to knowledge utilisation. African Governments would therefore do well to promote national statistics, including survey microdata, as an inexhaustible resource, which increases in value with re-use. Policies should encourage and sustain organisations that bridge the gap between data producers (in NSOs and academia) and data users (government decision-makers and researchers) to cultivate dialogue between these groups. Examples here include the AERC, which acts as an interface between researchers and policymakers in Africa (Hoffman, 1995:1-10).

## CONCLUSION AND RECOMMENDATIONS

SDAs can also span this divide, as well as play a more general advocacy role. Regular public communication via the media can assist government statisticians to advocate for the importance of statistics to both policymakers and the general public. This has been done successfully in South Africa, where the Statistician General presented a weekly column on statistical issues in a national daily newspaper. This served as an important tool for communication with data users and for dealing with controversial issues around statistics (Lehohla, 2005:61).

Data curation advocacy could include the introduction of incentives for the free sharing of research data. This can be done at a government level by allocating the necessary resources to NSOS and inculcating positive attitudes to data sharing among senior NSO staff. Requirements for access need to be put in place with regard to publicly funded survey research projects. Certainly, as in other parts of the world, funding can be linked to agreements for the eventual public availability of research data. At an institutional level, incentives will need to counteract the attractions of exclusive right to the data, and the right to publish during long embargo periods. Academic awards for the creation and timely sharing of quality social survey data could be introduced by universities to promote national data sharing. Compulsory access conditions could also be placed on African researchers to support data sharing. Such initiatives already exist in South Africa. For example, the University of Pretoria requires the depositing of research findings in its online institutional repository, UPSpace, as a precondition to the awarding of postgraduate degrees (Page-Shipp, 2007:14).

### **8.2.2 Policies for Funding of Data Curation**

Data curation policies could ensure the provision and stability of adequate funding to affect data management for growth in African countries. Data curation as a means of husbanding national information assets could be viewed by African policy-makers as a cost-effective development strategy and allocated a regular budget as a component of development funding. Data curation supports higher education and research and this activity could therefore be funded from budget allocations for education or research and development. Cash-strapped African governments could look to international grant-makers as a possible source of funding in this regard.

### **8.2.3 Counteracting Bureaucratic Constraints**

Policies are required to overcome bureaucratic hurdles with regard to data sharing, evidenced by the findings of the microdata access survey undertaken for this research (Appendix D). This study has shown that even where donor organisations have provided the necessary funding, without the adoption by African leaders of the principles of knowledge utilisation data sharing in African countries will not take place. This is because most data production in Africa occurs within official structures and much of the sharing of data would need to be undertaken by NSO staff and staff of other national data producing institutions such as central banks. The bureaucratic nature of government institutions in Africa means data sharing initiatives will not originate in this sphere without the cooperation and commitment of Statistical Directors who themselves will need authority invested in them by government ministers to make their data available for research purposes.

### **8.2.4 Policies to Advance Local Capacities for Data Curation**

A precondition for the development of effective data curation systems is an educated and skilled population. As shown in chapter eight of this study, the shortage of skills for data archiving is an obstacle to effective data usage in African countries. Policies aimed at knowledge utilisation for growth could support the allocation of resources to advance both technical and intellectual capacities in Africa for data curation. Professional associations of data managers could also benefit from official support to encourage a vibrant community of practice in the field.

Types of training which would advance skills for data curation include training in ICTs, statistical education, including training in data analysis, and data management training. Education departments in African countries could develop existing curricula and facilities at local universities to introduce new courses relevant to the subject. The recent nature of data curation as a research support activity in Africa means innovative strategies for effective data curation training programmes may need to be adopted. Training syllabi could aim to produce graduates who are adaptable as this is vital in an environment with fast-changing technologies. Statisticians need to be trained specifically in competencies related to government statistics and university courses could attempt to combine the roles of statistician and planner to produce policy-makers who can appreciate the value of empirical data for planning.

## CONCLUSION AND RECOMMENDATIONS

Governments in the region could also collaborate to utilise regional STCs as a resource for data management training. These are already established although they would need funding support to include data management courses in their curricula. The advantage of these training institutions is that they reach across the continent because of their regional nature, and can therefore provide skills to a wider audience than many national training institutions. For example the EASTC in Tanzania has enrolled trainees from nineteen countries in its regular courses (Muba, 2006: 147-149).

### **8.2.5 Support for Professional Associations**

African governments could provide support for the establishment of national and regional associations of data managers to benefit data curation. As detailed in chapter six of this study, international donor organisations have supported the establishment of an African association of survey data professionals in the belief that such associations can advance the objectives of knowledge utilisation on the continent. AASDA, launched in 2008, could be an important agency for statistical advancement and could at least be partially supported by public funds. A robust professional association of data managers in Africa could provide practitioners with access to a pool of expertise and allow the sharing of scarce resources and collaboration with regard to regional data curation systems. Such an association could provide mutual encouragement with regard to the preservation and sharing of data. This community of practice could also form a “pressure group” to influence producers and funders to encourage and sustain data curation in the region.

### **8.2.6 Establishment of the Regulatory Infrastructure for Data Curation**

Data curation policies could facilitate the establishment of regulatory infrastructures in African countries which are necessary to support the curation of African survey statistics. This research has revealed that enforcement measures are necessary to deal with conflicting stakeholder interests with regard to data sharing, which need to be formally addressed with input from all role-players. Laws and guidelines on data curation could define concepts, clarify roles and enforce rights and obligations with regard to the management of the country’s data assets (James, 2001:156).

Data curation laws would need to be regularly updated, to accommodate changes in technology. Laws regarding accountabilities could negotiate data ownership rights and balance

## CONCLUSION AND RECOMMENDATIONS

commercial rights against the public's information rights and the custodial obligations of governments and public institutions. Laws would need to be harmonised across the region, to ensure enforceability internationally. Trade policies would also be relevant here, as legislation would have to take into account possible cross-border data exchange (Henrici, 2004:32-37). Regional organisations such as the AU and UNECA would be well positioned to take on the task of assisting African countries to draw up appropriate legislation.

Laws and guidelines could include legislation to enlarge the authority and autonomy of NSOs in African countries and allocate more resources to these institutions. The results of the African microdata access survey conducted as part of this study reveal bureaucratic constraints to data sharing which could be mitigated by NSOs being independent of political influences (African microdata access survey, Appendix D). As discussed in chapter 8 which examined restrictions on data access, the issue of data confidentiality would also need to be addressed to prevent this becoming an intractable obstacle to national and regional data sharing in Africa. African government policies in this regard could be concerned with removing misconceptions regarding this issue and in this way allaying the fears of statistical staff regarding data misuse. This can be achieved by the provision of clear guidelines for appropriate levels of disclosure and steps that need to be taken to ensure the confidential nature of the data shared. Such legislation could alter for the better the current ad hoc nature of data access policies of data producers and data archiving organisations in African countries.

### **8.2.7 Development of the Necessary Technological Infrastructure**

This study has emphasises technological readiness as a requirement for effective data curation. The microdata access survey undertaken for this research reveals that inadequate technological infrastructures in African countries can hinder effective data management and data sharing by African NSOs (Appendix D). While the positive aspects of digitising survey data resources for broader access in Africa are obvious, how this process is managed and presented can determine whether data sharing will be optimally advantageous for academic research and policy formulation in Africa.

### **8.2.8 Development of Institutional Infrastructures for Data Curation**

It is the contention of this study that in African countries the efficient management and effective sharing of data needs to be formalised and institutionalised in state systems to change the data

## CONCLUSION AND RECOMMENDATIONS

sharing culture in these countries. In answering questions around the institutions that can be created in African countries to facilitate data curation this research emphasizes the need to firstly strengthen the data preservation and data sharing role of NSOs in the region, and secondly establish other institutions to support this activity, such as SDAs. These archives represent one type of “boundary bridging” institution which could act as an effective interface between the various components of national data systems (Kiregyera, 2005:75-80) and the development of a national data archive in each African country could support the preservation and re-use of the continent’s data resources.

Currently much African survey data resides with overseas organisations. Well-meaning foreign organisations have undertaken preservation of African data, but have in the process engaged in a form of “data imperialism” by restricting the release of these datasets into the public domain in Africa (Moultrie, 2008). Granted, the restrictions placed on access to microdata files in these collections may originally have been required by data depositors such as African NSOs. However, as governments begin to see the value in the sharing of empirical data, this may change, and it is important for future data sharing in the region that any extant copies of data files are archived in African institutions. The digital nature of survey data also means proximity to original files is not an imperative. However, a network of African SDAs can provide the infrastructure necessary to ensure copies of African survey microdata files also exist in African institutions.

### 8.2.8.1 SDAs within NSOs

A survey of the literature undertaken for the historical overview of SDA development discussed in chapter two shows that in Western Europe and North America NSOs play only a supporting role in national and regional data sharing. These institutions are major producers of census and survey data, but also collect other statistical information, such as administrative records. They can be viewed as one of the many suppliers of survey data to SDAs in these countries. However NSOs are the main producers of survey information in countries with poorly developed research support structures. For example, in Lithuania the main holder of empirical data is the Lithuanian Department of Statistics (Hausstein & de Guchteneire, 2002:69). A similar case applies in African countries.

Government financial support of the kind received by the European Survey Archives is not available in African countries for the development of a counterpart network in Africa. The

## CONCLUSION AND RECOMMENDATIONS

wealthy institutions of higher education that support SDAs in North America are scarce on the African continent. It may therefore be practical for a network of African SDAs to be established, at least initially, within African NSOs, in order to take advantage of the public funding and international donor support these institutions receive. This would ensure official sanction for data curation and enable the utilisation of the admittedly limited national technological infrastructures.

Situating African SDAs in NSOs presents a practical solution as this is where both the surveys and the survey management infrastructures exist in African countries. The establishment of SDAs within African NSOs would allow these archives to make use of the existing communication and administrative infrastructures of these organisations. Such data archives would also be in a position to take advantage of regional collaborations. For example a key principle in the African Charter on Statistics is co-ordination and co-operation between statistical organisations and establishing a network of SDAs could be a beneficial initial cooperative venture of NSO Directors in the region (Lehohla, 2007b).

SDAs within NSOs could make use of existing data management structures within these agencies, and could lead to improvements in official data management practices, as focus would be brought to bear on this aspect of official statistics. However, there are many obstacles to changing the conservative culture in African NSOs, which translates into an over-cautious attitude to the sharing of national data for research purposes. Situating SDA in NSOs could circumvent any independent quality control of the data produced. Establishing SDAs as part of official structures may also lead to the perception among researchers that the data produced has a government bias. This is particularly valid in an African context, with a history of intolerance between government and academia (Mwase, 1986:145).

### **8.2.8.2 SDAs within African Universities**

*The National Research and Development Strategy identifies the need to create 'centres and networks of excellence' in science and technology, including in the social sciences, as a key component of the human capital and transformation dimensions of government policy. It is envisaged that such centres will stimulate sustained distinction in research while simultaneously generating highly qualified human resource capacity in order to impact meaningfully on key national and global areas of knowledge (South Africa. Department of Science and Technology [DST], 2009).*

Universities, too have the infrastructure and commitment to sustain digital preservation (National Library of Australia, 2003:50). The advantages of situating SDAs within university

## CONCLUSION AND RECOMMENDATIONS

structures include the ease of access this would provide for a key user group – academics. Universities house both existing and potential users, as students may become future government policymakers. Archives based at universities could also assist communication between academics and national data producers and ensure data suppliers are aware of researchers' needs. At institutions of higher learning there is awareness of the value of data sharing, and the need for the long-term preservation of data for future secondary use. Thus academics would be keen to support university based SDAs. Issues around legitimacy could also be diminished if universities were responsible for data archives in African countries.

Funding issues would need to be addressed if SDAs are to be situated in African universities, rather than within central government structures. The primary task of universities is teaching and research, but in Africa financial resources for research are scarce, with the result that teaching often becomes the only financially supported function of these institutions. Most African universities spend up to 80 percent of their budget on teaching staff, and therefore have limited resources for maintaining and improving their research infrastructure, including supporting data sharing facilities (Task Force on Higher Education and Society, 2000:25). However, with funding from government sources and the international donor community SDAs at African universities could become “centres of excellence” and possibly attract international researchers keen to investigate African conditions “in situ”, whose work could promote these university based data archives and bring them to the attention of international grant-makers to assist them to access further funding.

### **8.2.8.3 SDAs as Autonomous Government Institutions**

Both NSOs and universities in Africa have meagre financial resources. NSOs may suffer from a lack of legitimacy and their conservative bureaucratic culture may make acceptance of and adaptation to a data sharing ethos difficult. Situating national SDAs within NSS but outside NSO structures could circumvent these obstacles to the production and dissemination of high-quality African data. This is where funding resources are most likely to be made available, either as direct contributions from government, or as donor funding channelled through national programmes. SADA is a good example of what can be accomplished in this regard. This archive, the only national SDA in Africa, is housed within the National Research Foundation, and funded as a unit of this research support institution (South African Data Archive, 2009).

Basing this type of facility within government structures but apart from NSOs could assist in the coordination of data curation in other government agencies. The existence of national data curating institutions could raise awareness within government departments of the need for sound data curation practices, and the successful operation of such data centres could influence attitude changes among government functionaries towards the preservation of national statistics for future research and policymaking. This could help overcome the conservative traditions of many African official institutions and encourage efficient systems for data archiving and data sharing for economic and social development in Africa.

### 8.3 Suggestions for future research

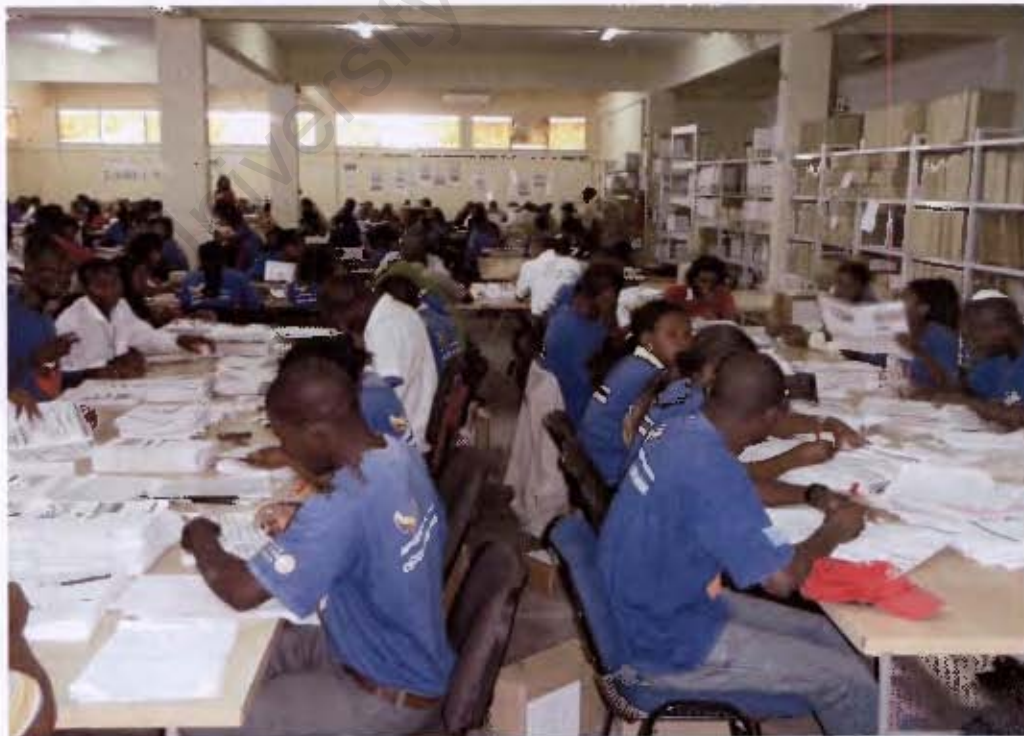
The findings with regard to barriers to effective data curation in African countries suggest that these need to be examined in more detail in order to mitigate some of these obstacles. It is suggested that further research should be concerned with the manner in which these obstacles could be overcome at a country level. This includes research into the provision of training for data management in African countries. Questions that need addressing in this regard should related to the type of training programmes that can be created to provide a reasonable percentage of Africans with the skills to both manage and use empirical data. This workforce could be utilised in government and academia to assist in changing the culture of data usage in the region.

The second area of interest for further examination is what strategies could be adopted with regard to data advocacy on the continent. Effective strategies in this regard would change the current situation in which donor organisations rather than African governments are usually the driving force in promoting data utilisation in the region. As this research as indicated, African leaders commitment to data usage is a prerequisite for effective data curation in countries of the region. For this reason options for the promotion of this activity to incumbent leaders in African countries need to be investigated in future research on the topic.

## CONCLUSION AND RECOMMENDATIONS



Photograph 13: DataFirst staff (including the author) and World Bank Consultants visiting the Census 2007 Warehouse in Mozambique in 2008. (Photo: Geoffrey Greenwell, 2008).



Photograph 14: Processing the 2007 Mozambican Census Results. (Photo: Geoffrey Greenwell, 2008).

## REFERENCES

### REFERENCES

- 3rd Africa Symposium on Statistical Development [ASSD]. 2007. [Presentations]. In 3<sup>rd</sup> Africa Symposium on Statistical Development, "Best Practice & Exchanging Experience: Africa 2010 Round of Population and Housing Censuses" Accra, Ghana 3-7 December, 2007. [Online]. Available: <http://www.paris21.org/documents/2790.pdf> (accessed 2009, September 30).
- Achikbache, B., Belkindas, M., Dinc, M. Eele, G. and Swanson, E. 2002. Strengthening statistical systems. In *A Sourcebook for Poverty Reduction Strategies*, ed. J. Klugman, Washington, DC: International Bank for Reconstruction and Development and World Bank. 1, Chapter 5: 159-186. [Online]. Available: [http://siteresources.worldbank.org/INTPRS1/Resources/383606-1205334112622/5457\\_chap5.pdf](http://siteresources.worldbank.org/INTPRS1/Resources/383606-1205334112622/5457_chap5.pdf) (accessed 2009, August 30).
- Adams, A. and Anum, J.D. 2005. Statistical information resources and development planning in Ghana. *Information Development* 21(2):138-144.
- Ademola, A. and Adeniran, A. 1998. *Study of data collection procedures*. Nairobi: African Economic Research Consortium.
- Africa Symposium on Statistical Development [ASSD]. 2006a. Report. In 2006 Africa Symposium on Statistical Development, "The 2010 Round of Population and Housing Censuses", Cape Town, South Africa, 30 January – 2 February, 2006. Document ASSD2006/01. [Online]. Available: <http://www.statssa.gov.za/asc/WebsiteReports/ASSD2006-01.pdf> (accessed 2009, August 30).
- Africa Symposium on Statistical Development [ASSD]. 2006b. Resolutions. In 2006 Africa Symposium on Statistical Development, "The 2010 Round of Population and Housing Censuses", Cape Town, South Africa, 30 January – 2 February, 2006. Document ASSD2006/03. [Online]. Available: <http://www.statssa.gov.za/asc/WebsiteReports/ASSD2006-03.pdf> (accessed 2009, August 30).

## REFERENCES

- African Association of Statistical Data Archivists. 2008. Minutes of the Inaugural Meeting, Cape Town, University of Cape Town, Breakwater Campus, 10-12 April 2008.
- African Capacity Building Foundation [ACBF]. 2004. The African Capacity Building Foundation and statistical capacity building in Africa. In Forum on African Statistical Development (FASDEV), Addis Ababa, Ethiopia, 12-13 May. [Online]. Available: [http://www.uneca.org/espdp/statistics/docs/fasdev\\_2004\\_e23.pdf](http://www.uneca.org/espdp/statistics/docs/fasdev_2004_e23.pdf) (accessed 2009, August 30).
- African Capacity Building Foundation [ACBF]. 2008. Reactivating the ASA [African Statistical Association] as a knowledge sharing platform. Harare: ACBF. [Online]. Available: [http://www.acbf-pact.org/whatsnew/docs/African\\_Statistical\\_Association.pdf](http://www.acbf-pact.org/whatsnew/docs/African_Statistical_Association.pdf) (accessed 2009, August 31).
- African Census Analysis Project [ACAP]. 2009. Census data. [Online]. Available: <http://www.acap.upenn.edu/census/> (accessed 2009, August 31).
- African Development Bank Group [ADB]. 2009. Knowledge: African Statistical Journal. [Online].. Available: <http://www.afdb.org/en/knowledge/statistics/publications/african-statistical-journal> (accessed 2009, September 30).
- African Economic Research Consortium [AERC]. 2009. Overview. [Online]. Available: <http://www.aercafrica.org/about/index.asp> (accessed 2009, August 30).
- African Regional Symposium on Telematics for Development, Ethiopia, 3-7 April '95. 1995. [Online]. Available: <http://www.africa.upenn.edu/Padis/telematics.html> (accessed 2009, August 30).
- African undersea cables. 2009. [Map]. [Online]. Available: <http://manypossibilities.net/african-undersea-cables/> (accessed 2009, August 30).

## REFERENCES

- African Union. 2009. African Charter on Statistics. Addis Ababa: Assembly of the African Union, 4 February. [Online]. Available: [http://www.africa-union.org/root/AU/Documents/Treaties/text/Charter\\_on\\_statistics%20-%20EN.pdf](http://www.africa-union.org/root/AU/Documents/Treaties/text/Charter_on_statistics%20-%20EN.pdf) (accessed 2009, August 30).
- African Union. Economic Affairs Department, Statistics Division. 2008. Aide mémoire. Experts meeting, "Establishing effective coordination mechanisms and enhancing financial contributions for statistical activities in Africa", Abidjan, Côte d'Ivoire, 4-5 August 2008. [Online]. Available: [www.africa-union.org/root/UA/Conferences/2008/aout/EA/4-5aout/Concept%20note%20of%20the%20Meeting%20in%20Abidjan.doc](http://www.africa-union.org/root/UA/Conferences/2008/aout/EA/4-5aout/Concept%20note%20of%20the%20Meeting%20in%20Abidjan.doc) (accessed 2009, August 31).
- Afrobarometer. 2009. Survey countries. [Online]. Available: <http://www.afrobarometer.org/countries.newmap.html> (accessed 2009, September 30).
- Akinyosoye, V. 2008. Repositioning the national statistical systems of African countries within the framework of international best practices: the case of Nigeria. *African Statistical Journal* 6: 191-219.
- Alexander, W.E., Cady, J. and Gonzalez-Garcia, J. 2008. The General Data Dissemination System: what has been accomplished after 10 years and where do we go from here? In: Alexander, W.E. ed. *The IMF's data dissemination initiative after 10 years*, 53-91. Washington: International Monetary Fund.
- Association for Computing Machinery [ACM]. website, 2009. [Online] Available: <http://www.acm.org> (accessed 2009, August 30).
- Bailey, T. 2005. The KwaZulu-Natal Income Dynamics Study (KIDS). In: *The dynamics of knowledge production and utilisation: fifteen case studies*, edited by T. Bailey, 494-553. Stellenbosch: Centre for Research on Science and Technology, Stellenbosch University. (The Production and Utilisation of Knowledge in Higher Education Institutions in South Africa, Vol. 4).

## REFERENCES

- Balepa, M. 2006. AFRISTAT, an example of cooperation and statistical partnership in Sub-Saharan Africa. *African Statistical Journal* 2: 135-143.
- Beaven, R. and Martin, R. 2004. Strengthening the quality and use of data in Africa. London: The Commission for Africa. [Online]. Available: [http://www.commissionforafrica.org/english/report/background/beaven\\_background.pdf](http://www.commissionforafrica.org/english/report/background/beaven_background.pdf) (accessed 2009, August 30).
- Bergdahl, B. 1990. *IFLA's Programme Advancement of Librarianship in the Third World (ALP): a Proposal for the Future*. Stockholm: Swedish Library Association.
- Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. 2003. *Conference on Open Access to Knowledge in the Sciences and Humanities, 20 - 22 Oct, Berlin*. [Online]. Available: <http://oa.mpg.de/openaccess-berlin/berlindeclaration.html> (accessed 2009, August 30)
- Bernardo, Tomas, 2008. Interview with Lynn Woolfrey at the offices of the Instituto Nacional de Estatística, Mozambique, December 5th.
- Booker, W., Singh, P. and L. Savane. 1980. *Household survey experience in Africa*. Washington: World Bank. (Living Standards Measurement Study Working Paper No. 6).
- Boruch, R. F. 1985. Definitions, products and distinctions in data sharing. In *Sharing research data*, edited by E. Fienberg, M. E. Martin, and L. M. L. Straf, 89-122. Washington: National Academies Press. [Online]. Available: <http://www.nap.edu/catalog/2033.html> (accessed 2009, August 31).
- Brief on the African Regional Technical Assistance Centers' (AFRITACs) involvement in statistical capacity building. 2004. *Forum on African Statistical Development (FASDEV)*, Addis Ababa, Ethiopia, 12-13 May. [Online]. Available: [http://www.uneca.org/espd/statistics/docs/fasdev\\_2004\\_e19.pdf](http://www.uneca.org/espd/statistics/docs/fasdev_2004_e19.pdf) (accessed 2009, August 30).

## REFERENCES

- Bugliarello, G. 1995. The global generation, transmission and diffusion of knowledge: how can developing countries benefit? In National Research Council, Office of International Affairs, Technology and Development Steering Committee, and World Bank, Finance and Private Sector Development. 1995. *Marshalling technology for development: proceedings of a symposium, 28-30 November 1994*, Arnold and Mabel Beckman Center, Irvine, California. Washington: National Academies Press, 61-82. [Online]. Available [http://books.nap.edu/openbook.php?record\\_id=5022](http://books.nap.edu/openbook.php?record_id=5022) (accessed 2009, August 31).
- Cairo Demographic Center. [2009?]. Short training programs (Training courses). [Online]. Available: <http://www.cdc-egypt.org/courses.asp?mymain=4&mysub=&SubSubID=15> (accessed 2009, September 15).
- Carley, M. and Card, J. 2000. The Social Science Electronic Data Library: serving the needs of data librarians and users. *IASSIST Quarterly* 24(3): 8-14.
- Center for Regulatory Effectiveness. [2002?]. Data quality OMB to begin implementing new data quality law. Washington: Center for Regulatory Effectiveness. [Online] Available: [http://www.thecre.com/quality/OMB\\_Implements\\_New\\_DataQualityLaw.html](http://www.thecre.com/quality/OMB_Implements_New_DataQualityLaw.html) (accessed 2009, August 30).
- Central Statistical Authority of Ethiopia (CSA). 2008. Home page. [Online]. Available: <http://www.csa.gov.et/index.htm> (accessed 2009, September 15).
- The CESSDA ESFRI Project: setting up a one-stop shop for European data. 2008. Concurrent Session H1 at the IASSIST 2008 Conference: technology of Data: collection, communication, access and preservation, Stanford University, Palo Alto, California, 27-30 May. [Online]. Available: <https://www.stanford.edu/group/ADS/cgi-bin/drupal/session/the-cessda-esfri-project-setting-up-a-onestop-shop-european-data> (accessed 5 September 2009).

## REFERENCES

- Chander, R., Grootaert, C. and G. Pyatt. 1980. *Living standards surveys in developing countries*. Washington: The World Bank. (Living Standards Measurement Study Working Paper No. 1).
- The Chinese University of Hong Kong. {2009?}. Databank for China Studies. [Online]. Available: <http://www.usc.cuhk.edu.hk/databank.asp> (accessed 2009, September 30).
- Ching'anda, E. F. and Ntozi, J.P.M. 1998. Training employees for African statistical offices. *International statistical review* 66(2; August): 235-244.
- Ching'anda, E.F. 1999. Towards improving official statistics in Africa. Addis Ababa: United Nations Economic Commission for Africa. [Online]. Available: <http://www.stat.fi/isi99/proceedings/arkisto/varasto/chin0829.pdf> (accessed 2009, July 29).
- Clubb, J. M., Austin, E. W. Geda, C. L. and Traugott, M. W. 1985. Sharing research data in the social sciences. In Fienberg, E. Martin, M.E. and L. M. L. Straf, Editors. *Sharing research data*, 39-88. Washington: National Academies Press. [Online]. Available: <http://www.nap.edu/catalog/2033.html> (accessed 2009, August 31).
- Committee on Issues in the Transborder Flow of Scientific Data and National Research Council. 1997. *Bits of power: issues in global access to scientific data*. Washington, DC: National Academies Press. [Online]. Available: [http://cart.nap.edu/cart/pdfaccess.cgi?&record\\_id=5504](http://cart.nap.edu/cart/pdfaccess.cgi?&record_id=5504) (accessed 2009, August 31).
- Committee on the Geographic Foundation for Agency 21, Committee on Geography, Mapping Science Committee and the National Research Council. 2002. *Down to Earth: geographical information for sustainable development in Africa*. Washington: National Academies Press. [Online]. Available: [http://www.nap.edu/catalog.php?record\\_id=10455](http://www.nap.edu/catalog.php?record_id=10455) (accessed 2009, August 31).
- Compact Oxford English dictionary. 2009. [Online]. Available: <http://www.askoxford.com> (accessed 2009, August 30).

## REFERENCES

- Cook, L. 2003. Summary of discussants' main points. *Statistical confidentiality and access to microdata: proceedings of the Seminar Session of the 2003 Conference of European Statisticians*, Geneva, 10-12 June 2003, 11-20. New York: United Nations Economic Commission for Europe. [Online]. Available: <http://www.unece.org/stats/publications/statistical.confidentiality.pdf> (accessed 2009, August 30).
- Cooper, A. 2005. A proposed methodology and infrastructure for standards development and implementation within a national statistical agency. *The Fourteenth Conference of Commonwealth Statisticians*, Cape Town, 5-9 September, 2005. [Online]. Available: [http://www.statssa.gov.za/commonwealth/presentations/A\\_Cooper.pdf](http://www.statssa.gov.za/commonwealth/presentations/A_Cooper.pdf) (accessed 2009, August 30).
- Council of European Social Science Data Archives [CESSDA]. 2008. CESSDA PPP: Preparatory Phase Project for a Major Upgrade of the Council of European Social Science Data Archives. Research Infrastructure. [Online]. Available: <http://www.nsd.uib.no/cessda> (accessed 2009, August 30).
- Da Cruz, F. [2009?]. Columbia University Computing History: A Chronology of Computing at Columbia University. Columbia University Information Technology. [Online]. Available: <http://www.columbia.edu/acis/history> (accessed 2009, September 26).
- Data Documentation Initiative [DDI], 2009. DDI 3.0. [Online]. Available: <http://www.ddialliance.org/ddi3/index.html> (accessed 2009, August 30).
- DataFirst. 2009a. List of survey datasets. [Online]. Available: <http://www.datafirst.uct.ac.za/home/> (accessed 2009, August 30).
- DataFirst. 2009b. Data Quality Project. [Online]. Available: <http://www.datafirst.uct.ac.za/home/index.php?/data-quality-project> (accessed 2009, August 30).
- DataFirst. 2009c. UCT Summer Training Programme in Social Science Using Survey Data. [Online]. Available: [http://www.datafirst.uct.ac.za/saproject/sa\\_project/index.html](http://www.datafirst.uct.ac.za/saproject/sa_project/index.html) (accessed 2009, September 15).

## REFERENCES

- de Gier, E., Henke, R and J. Vilgen. 2004. "Knowledge utilisation" in Dutch social policy: the case of the Disability Insurance Act (WAO). *International Social Science Journal* 56 (179):17-34.
- de Vaus, D. 2002. *Surveys in social research*. 5th Edition. London: Routledge.
- Demographic and health surveys. [DHS]. 2009. About DHS. [Online]. Available: <http://www.measuredhs.com/aboutdhs/history.cfm> (accessed 2009, September 30).
- Digital Curation Centre [DCC]. 2005. About the DCC. [website]. Edinburgh: DCC. Available: <http://www.dcc.ac.uk/FAQs/data-curator> (accessed 2009, August 30).
- Dupriez, O. 2008a. Establishing data archives in developing countries: some initial steps. IASSIST 2008 conference: technology of data: collection, communication, access and preservation, Stanford University, Palo Alto, 26-30 May, 2008. [Online]. Available: <https://www.stanford.edu/group/ADS/cgi-bin/drupal/establishing-data-archives-developing-countries-some-initial-steps> (accessed 2009, August 31)
- Dupriez, O. 2008b. Personal e-mail to Lynn Woolfrey (15 August).
- Dunne, S. and Austin, E.W. 1998. Protecting confidentiality in archival data resources. *ICPSR Bulletin* 19(1):1-6.
- Durant, V.L. and Menken, J. Eds. 2002. *Leveraging longitudinal data in developing countries: report of a workshop*. Washington: National Research Council.
- Dutta, S. and Mia, I., eds. 2009. *The global Information Technology Report 2008-2009: mobility in a networked world*. Geneva: World Economic Forum. [Online]. Available: <http://www.insead.edu/v1/qitr/wef/main/fullreport/index.html> (accessed 2009 August 30).
- e-Strategies: best practices and lessons learnt. 2007. Addis Ababa: Economic Commission for Africa. [Online]. Available: <http://www.uneca.org/aisi/nici/Documents/NICI-Book.pdf> (accessed 2009, August 31).

## REFERENCES

- East Africa gets high-speed web. 2009. *BBC News*. [Online]. Available: <http://news.bbc.co.uk/2/hi/africa/8165077.stm> (accessed 2009, August 30).
- Eastern Africa Statistical Training Centre [EASTC]. 2006. EASTC teaching programmes. [Online]. Available: [http://www.eastc.ac.tz/index.php?module=pagemaster&PAGE\\_user\\_op=view\\_page&PAGE\\_id=6&MMN\\_position=11:11](http://www.eastc.ac.tz/index.php?module=pagemaster&PAGE_user_op=view_page&PAGE_id=6&MMN_position=11:11) (accessed 2009, September 20).
- Economic and Social Data Service [ESDS]. Rural Economy and Land Use Programme Data Support Service. 2006. *Guidance on data management*. Colchester, Essex: UK Data Archive. [Online]. Available: <http://www.esds.ac.uk/news/publications/reluaug2006.pdf> (accessed 2009, August 30).
- Economic and Social Data Service website. [ESDS] 2009. [Online]. Available: <http://www.esds.ac.uk> (accessed 2009, August 30).
- Economic and Statistical Observatory of Sub-Saharan Africa [AFRISTAT]. 2009. Présentation. [Online]. Available <http://www.afristat.org/?pg=100#afristat> (accessed 2009, August 30).
- Eele, G. 1989. The organisation and management of statistical services in Africa: why do they fail? *World Development* 17(3):431-438.
- Eele, G., and Chinganya, O. 2005. Using the GDDS to build statistical capacity in Africa. *African Statistical Journal* 1:18-29.
- Esanu, J.M. and Uhler, P.F. Eds. 2004. *Open access and the public domain in digital data and information for science*. Washington: National Academies Press. [Online]. Available: [http://books.nap.edu/openbook.php?record\\_id=11030&page=69](http://books.nap.edu/openbook.php?record_id=11030&page=69) [accessed 2009, August 31).
- European Commission. EUROSTAT. 2009a. Introduction. [Online]. Available: [http://epp.eurostat.ec.europa.eu/portal/page/portal/about\\_eurostat/corporate/introduction](http://epp.eurostat.ec.europa.eu/portal/page/portal/about_eurostat/corporate/introduction) (accessed 2009, September 26).

## REFERENCES

- European Commission. EUROSTAT. 2009b. Key historical facts. [Online]. Available: [http://epp.eurostat.ec.europa.eu/portal/page/portal/about\\_eurostat/corporate/introduction/historical](http://epp.eurostat.ec.europa.eu/portal/page/portal/about_eurostat/corporate/introduction/historical) (accessed 2009, September 30).
- European Commission. EUROSTAT. 2009c. Welcome to the MEDSTAT programme. [Online]. Available: <http://epp.eurostat.ec.europa.eu/portal/page/portal/medstat/introduction> (accessed 2009, August 31).
- European Union. 1997. Official journal, L 052 (29 April): 1-7. Available: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31997R0322:EN:HTML> (accessed 2009, September 5).
- Fienberg, E., Martin M.E. and L.M.L. Straf, Eds. 1985. *Sharing research data*. Washington: National Academies Press. [Online]. Available: <http://www.nap.edu/catalog/2033.html> (accessed 2009, August 31).
- Financial tide goes out on incumbent telco privatizations. 2008. *Balancing Act News Update 422, 20 February*. [Online]. Available: [www.balancingact-africa.com/news/back/balancing-act\\_422.html](http://www.balancingact-africa.com/news/back/balancing-act_422.html) (accessed 2009, August 31).
- Foundation Conference for an International Data Forum. 2007. Conference report and next steps. *Foundation Conference for the International Data Forum, Beijing, 5<sup>th</sup> -7<sup>th</sup> June, 2007*. [Online]. Available: [http://www.esrcsocietytoday.ac.uk/esrcinfocentre/images/idf%20conference%20report%202007\\_tcm6-21126.pdf](http://www.esrcsocietytoday.ac.uk/esrcinfocentre/images/idf%20conference%20report%202007_tcm6-21126.pdf) (accessed 2009, August 31).
- Gervais, R.R. and Marcoux, R. 1993. Saving Francophone Africa's statistical past. *History in Africa* 20:385-390. [Online]. Available: <http://www.jstor.org/stable/3171984?origin=JSTOR-pdf> (accessed 2009, August 31).
- Gillwald, A., and Stork, C. 2008. Towards evidence-based ICT policy and regulation: ICT access and usage in Africa Volume one. Johannesburg: Research ICT Africa (RIA). [Online]. Available: <http://www.researchictafrica.net/new/images/uploads/ria-policy-paper-ict-access-and-usage-2008.pdf> [accessed 2009, August 31].

## REFERENCES

- Gray, E. 2007. A policy workshop on access to data. *Gray area*. [Weblog]. Available: <http://blogs.uct.ac.za/blog/gray-area/2007/10/09/a-policy-workshop-on-access-to-data> (accessed 2009, August 31).
- Greenwell, G. 2008. Personal e-mail to Lynn Woolfrey (18 December).
- Hauben, M. 2009. *Behind the Net - The untold history of the ARPANET*. Porto: Instituto Superior de Engenharia do Porto [ISEP], Departamento de Engenharia Informática. [Online]. Available: <http://www.dei.isep.ipp.pt/docs/arpa> (accessed 2009, July 30).
- Hausstein, B. and de Guchteneire, P. 2002. *Social Science Data Archives in Eastern Europe: results, potentials and prospects of the archival development*. Bergisen Gladbach: E. Ferger Verlag.
- Hedrick, T. E. 1985. Justification for and obstacles to data sharing. In *Sharing research data*, edited by E Fienberg, M. E. Martin and L. M. L. Straf, 123-147. Washington: National Academies Press. [Online]. Available: <http://www.nap.edu/catalog/2033.html> (accessed 2009, August 31).
- Hendrici, I. 2004. Approach to the formulation of a National Information Policy for South Africa. *South African journal of librarianship and information science* 70 (1):30-38.
- Higher education in developing countries: peril and promise*. 2000. Washington: World Bank.
- History of Computing Project. 2009. [Online]. Available <http://www.thocp.net> (accessed 2009, September 26).
- Hoffman, J. 1995. *Linking economic research and policy-making: the role of the AERC*. Nairobi: African Economic Research Consortium.
- Humphrey, C. K., Estabrooks, A. A. Norris, J. R., Smith, J. E. and Hesketh, K. L. 2000. Archivist on board: contributions to the research team. *Forum: Qualitative social research* (3):1-4. [Online] Available: <http://www.qualitative-research.net/index.php/fqs/article/view/1022/2204> (accessed 2009, August 30).

## REFERENCES

- If you can't measure it, you can't manage it. 2007. *Cape Times Business Report*, October 4.
- Inter-university Consortium for Political and Social Research (ICPSR). 2009. Guide to social science data preparation and archiving: best practice throughout the data life cycle. 4th ed. Ann Arbor: ICPSR. [Online]. Available: <http://www.icpsr.umich.edu/icpsrweb/ICPSR/access/dataprep.pdf> (accessed 2009, August 31).
- Inter-University Consortium for Political and Social Research [ICPSR]. [2009?]a. [Online]. Available: <http://www.icpsr.umich.edu/ICPSR/> (accessed 2009, August 30).
- Inter-University Consortium for Political and Social Research [ICPSR]. [2009?]b. Membership. [website]. Ann Arbor: ICPSR. [Online]. Available: <http://www.icpsr.umich.edu/ICPSR/membership/index.html> (accessed 2009, August 30).
- Inter-University Consortium for Political Research [ICPSR]. Archive. 1976. Memo to ICPSR official representatives regarding updates and corrections to data descriptions and documentation. *ICPSR Bulletin*, no. 1 (16 January). [Online]. Available: <http://www.icpsr.umich.edu/files/ICPSR/org/publications/bulletin/1976-01.pdf> (accessed 2009, August 30).
- International Association for Social Science Information Services and Technology [IASSIST]. 2009. About IASSIST. [Online]. Available: <http://www.iassistdata.org> (accessed 2009, August 30).
- International Committee for Social Sciences Documentation: ten years of activity. 1962. *International Social Science Journal* 14(1): 177-191.
- International Council for Science. Committee on Data for Science and Technology [CODATA]. Preservation of and Access to Scientific and Technical Data in Developing Countries. [Online]. Available: <http://www.codata.org/taskgroups/TGpreservation/index.html> (accessed 2009, August 31).

## REFERENCES

- International Federation of Data Organizations for the Social Science [IFDO]. 2005. International networks and data archives. [web site] Available: <http://www.ifdo.org/network/index.html> (accessed 2009, August 30).
- International Household Survey Network. 2006. IHSN central survey catalog. [Online]. Available: <http://www.surveynetwork.org/home/?lvl1=activities&lvl2=catalog&lvl3=surveys> (accessed 2009, September 15).
- International Household Survey Network. [2008?]. Accelerated Data Program. [Online]. Available: <http://www.ihsn.org/adp/> (accessed 2009, September 15).
- International Labour Office [sic] [ILO]. Policy Integration Office. Bureau of Statistics. Statistical activities of the ILO. In Forum on African Statistical Development [FASDEV], Addis Ababa, Ethiopia, 12-13 May 2004. [Online]. Available: [http://www.uneca.org/espd/statistics/docs/fasdev\\_2004\\_e26.pdf](http://www.uneca.org/espd/statistics/docs/fasdev_2004_e26.pdf) (accessed 2009, September 8).
- International Monetary Fund [IMF]. 2002. Botswana: Report on the Observance of Standards and Codes – data module, response by the authorities, and detailed assessment using the Data Quality Assessment Framework. Washington, D.C.: IMF. IMF Country report no 04/87. [Online]. Available: <http://www.imf.org/external/pubs/ft/scr/2002/cr0283.pdf> (accessed 2009, August 30).
- International Monetary Fund [IMF]. 2004. *Burkina Faso: Report on the Observance of Standards and Codes – data module, response by the authorities, and detailed assessment using the Data Quality Assessment Framework*. Washington, D.C.: IMF. IMF Country report no 04/87. [Online]. Available: <http://www.imf.org/external/pubs/ft/scr/2004/cr0487.pdf> (accessed 2009, August 30).

## REFERENCES

- International Monetary Fund [IMF]. 2007. *Chad: Report on the Observance of Standards and Codes – data module, response by the authorities, and detailed assessment using the Data Quality Assessment Framework*. Washington, D.C.: IMF. IMF Country report no 07/300. [Online]. Available: <http://www.imf.org/external/pubs/ft/scr/2007/cr07300.pdf> (accessed 2009, August 30).
- International Monetary Fund. 2009. Dissemination Standards Bulletin Board [DSBB]. [Online]. Available: <http://dsbb.imf.org/Applications/web/dsbbhome/> (accessed 2009, September 25).
- International Monetary Fund. East Africa Regional Technical Assistance Centre (AFRITAC). 2004. A brief on the Africa Regional Technical Assistance Centers' (AFRITACS) involvement in statistical capacity building. In: *Meeting of the Forum on Africa Statistical Development (FASDEV)*, Addis Ababa, 12-13 May 2004. [Online]. Available: [http://www.uneca.org/espd/statistics/docs/fasdev\\_2004\\_e19.pdf](http://www.uneca.org/espd/statistics/docs/fasdev_2004_e19.pdf) (accessed 2009, September 15).
- Internet Systems Consortium Internet Host Counts. 2009. [Online]. Available: <http://ftp.isc.org/www/survey/reports/current/report.bynum> (accessed 2009, August 31)
- James, T. ed. 2001. *An information policy handbook for Southern Africa*. Ottawa: International Development Research Centre (IDRC). [Online] Available: [http://www.idrc.ca/fr/ev-11402-201-1-DO\\_TOPIC.html](http://www.idrc.ca/fr/ev-11402-201-1-DO_TOPIC.html) (accessed 2009, August 31).
- Johnson, C. 2005. Improving Public Access to and Dissemination of Government Information and Using the Federal Enterprise Architecture Data Reference Model. Memorandum for the Heads of Executive Departments and Agencies. Washington: Executive Office of the President. [Online]. Available: <http://www.whitehouse.gov/omb/memoranda/fy2006/m06-02.pdf> (accessed 2009, August 30).
- Kalton, G. 2000. Developments in survey research in the past 25 years. *Survey Methodology* 26(1):3-10. Ottawa: Statistics Canada.

## REFERENCES

- King, J. 2003. Recent EU legislation for research access to confidential data: implementation and implications. In *Statistical confidentiality and access to microdata: proceedings of the Seminar Session of the 2003 Conference of European Statisticians*. Geneva, 10-12 June 2003, 97-106. [Online]. Available: <http://www.unece.org/stats/publications/statistical.confidentiality.pdf> (accessed 2009, August 31).
- Kiregyera, B. 2001. Novel statistical capacity building initiatives: Addis Ababa Plan of Action and PARIS21. In 53rd Session of the International Statistical Institute, Seoul, Korea, 22-29 August 2001. [Online] Available: <http://isi.cbs.nl/iamamember/CD2/pdf/881.PDF> (accessed 2009, August 30).
- Kiregyera, B. 2005. A case and some actions for improving statistical advocacy in poor developing countries. *African Statistical Journal* 1: 70-84.
- Kiregyera, B. 2007. Statistics for managing for results: challenges, new initiatives and prospects for improving statistical systems in Africa. *African Statistical Journal* 5: 135-160.
- Kirkman, G.S., Cornelius, P.K., Sachs, J.D. and Schwab, K. 2002. *The global information technology report 2001-2002*. New York: Oxford University Press.
- Lancetti, M. Statistical cooperation of the European Commission [EUROSTAT]. 2004. In Forum on African Statistical Development [FASDEV], Addis Ababa, Ethiopia, 12-13 May 2004. [PowerPoint presentation]. Available: [http://www.uneca.org/espd/statistics/docs/fasdev\\_2004\\_e22.ppt](http://www.uneca.org/espd/statistics/docs/fasdev_2004_e22.ppt) (accessed 2009, September 8).
- Landais, J. ([Jean-Marc.LANDAIS@oecd.org](mailto:Jean-Marc.LANDAIS@oecd.org)) 2008. NSDS status in IDA countries : updated progress report as of 30 September 2008. [Personal e-mail, 22 October] to Lynn Woolfrey ([lynn.woolfrey@uct.ac.za](mailto:lynn.woolfrey@uct.ac.za)).

## REFERENCES

- Lane, J. 2003. Uses of microdata: keynote speech. In *Statistical confidentiality and access to microdata: proceedings of the Seminar Session of the 2003 Conference of European Statisticians*, Geneva, 10-12 June 2003, 11-20. New York: United Nations Economic Commission for Europe. [Online]. Available: <http://www.unece.org/stats/publications/statistical.confidentiality.pdf> (accessed 2009, August 30).
- Lehohla, P. 2005. Statistics South Africa in transition: reflections on a decade of statistical practice (1994-2004). *African Statistical Journal* 1:48-69.
- Lehohla, P. 2007a. African charter paves way for professional data crunching across the continent. *Business Report*, 21 June. [Online]. Available: <http://www.busrep.co.za/index.php?fSectionId=2511&fArticleId=3894650> (accessed 2009, August 31)
- Lehohla, P. 2007b. Central data system allows more informed evaluation of accuracy. *Business Report* (14 June). [Online]. Available: <http://www.busrep.co.za/index.php?fSectionId=553&fArticleId=3882832> (accessed 2009, September 15).
- Lehohla, P. 2008. Statistical Development in Africa in the context of the Global Statistical System. In *Thirty-ninth session of the Statistical Commission Regional Statistical Development in Africa*, 26 - 29 February. [Online]. Available: <http://unstats.un.org/unsd/statcom/doc08/BG-AfricaStatDev.pdf> (accessed 2009, August 30).
- Lesaoana, M.A. 1997. Data archiving in Africa: the South African experience. *IASSIST Quarterly* 21(1):4-7.
- Lewis-Beck, M.S., Bryman, A. and T.F. Liao. Editors. 2004. *The Sage encyclopedia of social science research methods: Vols 1-3*. Thousand Oaks: Sage Publications.
- Loureiro, J.D. 2009. [Personal interview with Lynn Woolfrey, 18 August]. Durban. (Unpublished).

## REFERENCES

- Lubanski, A. 1996. Social science data services during the last five years of the millennium: developments in the delivery and support of data services for academic research in Europe and North America. *IASSIST Quarterly* 20(4):17-27.
- Lufumpa, C.L. and Mouyelo-Katoula, M. 2005. Strengthening statistical capacity in Africa under the framework of the International Comparisons Program for Africa (ICP-Africa). *African Statistical Journal* 1: 30-47.
- Lynn, P. 2004. Editorial: Measuring and communicating survey quality. *Journal of the Royal Statistical Society A*, 167(4): 575-578.
- Mackie, C. and Bradburn, N. Eds. 2000. *Improving access to and confidentiality of research data*. Washington: National Academy of Sciences.
- Manning, R. 2006. Keynote Speech. Forum on African Statistical Development (FASDev II), Addis Ababa, 6-10 February, 2006. [Online]. Available: [www.uneca.org/fasdev/speech\\_richard\\_manning.htm](http://www.uneca.org/fasdev/speech_richard_manning.htm) (accessed 2009, August 31).
- The Marrakesh Action Plan for Statistics: better data for better results: an action plan for improving development statistics. 2004. Discussion document presented to the Second International Roundtable on Managing for Development Results, Marrakesh, Morocco, February 4-5, 2004. [Online]. Available: <http://www.mfdr.org/2ndRoundtable.html> (accessed 2009, August 31)
- Mbock, C. G., Ngo-Mpeck, M.L., Kom, D. and Belinga, J.M.Z. 2004. Policy utilisation of research results in Cameroon. *International Social Science Journal* 56 (179): 37-45.
- MEDSTAT II. 2008. Arab Republic of Egypt country statistical situation report: version 2.0 (final). Luxembourg: EUROSTAT. [Online]. Available: [http://epp.eurostat.ec.europa.eu/portal/page/portal/medstat/documents/EGYPT%20CSSR%20FINAL%20VERSION%202008\\_12\\_2008%20\\_2\\_.pdf](http://epp.eurostat.ec.europa.eu/portal/page/portal/medstat/documents/EGYPT%20CSSR%20FINAL%20VERSION%202008_12_2008%20_2_.pdf) (accessed 2009, September 8).

## REFERENCES

- Meltzer, H. 2003. Harmonisation of concepts, definitions and classifications. United Nations Statistics Division Expert Group meeting on setting the scope of social statistics, New York, 6-9 May, 2003. [Online]. Available: [http://unstats.un.org/unsd/demographic/meetings/egm/Socialstat\\_0503/docs/no\\_20.pdf](http://unstats.un.org/unsd/demographic/meetings/egm/Socialstat_0503/docs/no_20.pdf) (accessed 2009, August 31).
- Mochmann, E. 2005. European cooperation in social science data dissemination. [Online]. Available: [http://www.ifdo.org/data/data\\_access\\_conditions.html](http://www.ifdo.org/data/data_access_conditions.html) (accessed 2009, August 31).
- Mochmann, E. and de Guchteneire, P. 2005. The social science data archive step-by-step. Cologne: International Federation of Data Organizations for the Social Sciences. Available: [http://www.ifdo.org/data/data\\_archive\\_workflow01.html](http://www.ifdo.org/data/data_archive_workflow01.html) (accessed 2009, August 31).
- Moultrie, T. 2008. Interview with Lynn Woolfrey at the Centre for Actuarial Research, University of Cape Town, Thursday 31 July.
- Mouyelo-Katoula, M. 2006. Rethinking statistics for national development in Africa. *African Statistical Journal* 2: 19-29.
- Muba, V.E. 2004. Assistance in statistical training: the experience of the Eastern Africa Statistical Training Centre. In Forum on African Statistical Development [FASDEV], Addis Ababa, Ethiopia, 12-13 May 2004. [Online]. Available: [http://www.uneca.org/espdp/statistics/docs/fasdev\\_2004\\_e12.pdf](http://www.uneca.org/espdp/statistics/docs/fasdev_2004_e12.pdf) (accessed 8 September 2009).
- Mudesir Seid, Y. 2006. Implementing a National Data Archive in Ethiopia: Challenges and Experiences. *Paper presented at the IASSIST Conference: Data in a World of Networked Knowledge*, 23-26 May, University of Michigan Ann Arbor, Michigan.
- Multilingual Access to Data Infrastructures of the European Research Area [MADIERA]. 2009. The MADIERA Project. [Online]. Available <http://www.madiera.net> (accessed 2009, August 30).

## REFERENCES

- Munoz, J. 2005. A guide for data management of household surveys. In *Household sample surveys in developing and transition countries*, 305-334. New York: United Nations. (Studies in methods, series F, no. 96). Available:  
[http://secgen.comunidadandina.org/ANDESTAD/gt/gt12/docs/household\\_surveys.pdf](http://secgen.comunidadandina.org/ANDESTAD/gt/gt12/docs/household_surveys.pdf) (accessed 2009, August 31).
- Musgrave, S. 2003. The metadata dynamic: ensuring data has a long and healthy life. In Banks, R., Currall, J., Francis, J., Gerrard, L., Khan, R., Macer, T., Rigg, M., Ross, E., Taylor, S. and Westlake, A. (Eds.),. *ASC 2003: the impact of new technology on the survey process: proceedings of the 4th ASC International Conference*. Berkeley: Association for Survey Computing. [Online]. Available:  
[http://www.essex.ac.uk/hhs/staff/musgrave/ascmetadata\(revised\).pdf](http://www.essex.ac.uk/hhs/staff/musgrave/ascmetadata(revised).pdf) (accessed 2009, August 31).
- Mutula, S.M. 2003. Cyber cafe industry in Africa. *Journal of Information Science* 29: 489-497.
- Mwase, N. 1986. Social science research in Eastern and Southern Africa. *International Social Science Journal* 38(1) 139-147.
- N'Guessan, K. and Chitou, B. 2006. ENSEA of Abidjan: in the heart of statistical training in Francophone Africa. *The African Statistical Journal* 2(May): 151-157. [Online]. Available:  
[http://www.afdb.org/fileadmin/uploads/afdb/Documents/Publications/African.Statistical.Journal\\_Vol2\\_2.Articles\\_9.ENSEA\\_Abidjan.pdf](http://www.afdb.org/fileadmin/uploads/afdb/Documents/Publications/African.Statistical.Journal_Vol2_2.Articles_9.ENSEA_Abidjan.pdf) (accessed 2009, September 8).
- Nassimbeni, M. 1998. The information society in South Africa: from global origins to local vision. *South African Journal of Library and Information Science* 66 (4): 154-161.
- National Library of Australia. 2003. Guidelines for the preservation of digital heritage. Paris: United Nations Educational, Scientific and Cultural Organisation. [Online] Available:  
<http://unesdoc.unesco.org/images/0013/001300/130071e.pdf> (accessed 2009, August 30).

## REFERENCES

- National Research Council, Office of International Affairs, Technology and Development Steering Committee, and World Bank, Finance and Private Sector Development. 1995. *Marshalling technology for development: proceedings of a symposium, 28-30 November 1994*, Arnold and Mabel Beckman Center, Irvine, California. Washington: National Academies Press. Available [http://books.nap.edu/openbook.php?record\\_id=5022](http://books.nap.edu/openbook.php?record_id=5022) (accessed 2009, August 31).
- National Research Foundation [NRF]. 2008. 1<sup>st</sup> African digital curation conference: programme. [Online]. Available: <http://stardata.nrf.ac.za/nadicc/index.html> (accessed 2009, September 30).
- National Research Foundation [NRF]. 2009. CODATA South African National Committee. [Online]. Available: <http://stardata.nrf.ac.za/Codata/CodataSouthAfrica.html> (accessed 2009, September 30).
- Neotel. 2009. Home page. [Online]. Available: <http://www.neotel.co.za> (accessed 2009, August 30).
- Networked Social Science Tools and Resources [NESSTAR]. 2009. NESSTAR: publish data on the web. [Online]. Available: <http://www.nesstar.com> (accessed 2009, August 30).
- New Partnership for Africa's Development short term action plan: infrastructure: executive summary [draft], 2002. [Online]. Available: [www.afdb.org/fileadmin/uploads/afdb/Documents/Project-and-Operations/00473839-EN-PAPER-NEPAD-INFRASTRUCTURE-SHORT-TERM-ACTION-PLAN-EXEC-SUMMARY.PDF](http://www.afdb.org/fileadmin/uploads/afdb/Documents/Project-and-Operations/00473839-EN-PAPER-NEPAD-INFRASTRUCTURE-SHORT-TERM-ACTION-PLAN-EXEC-SUMMARY.PDF) (accessed 2009, August 31).
- Niva, M. Sundgrun, B. and Lyberg, I. 2003. Statistical confidentiality and microdata. In *Statistical confidentiality and access to microdata: proceedings of the Seminar Session of the 2003 Conference of European Statisticians, 21-27*. Geneva, 10-12 June 2003. Available: <http://www.unece.org/stats/publications/statistical.confidentiality.pdf> (accessed 2009, August 31).

## REFERENCES

- Nwalo, K.I.N. 2000. Managing information for development in the 21st Century: prospects for African libraries, challenges to the world. *The 66th IFLA Council and General Conference, Jerusalem, Israel, 13-18 August*. [Online]. Available: <http://www.ifla.org.sg/IV/ifla66/papers/012-114e.htm> (accessed 2009, August 31).
- Nyanzi, T.M. 2004. Makerere University Institute of Statistics and Applied Economics (ISAE). In *Forum on African Statistical Development [FASDEV]*, Addis Ababa, Ethiopia, 12-13 May 2004. [Online]. Available: [http://www.uneca.org/espd/statistics/docs/fasdev\\_2004\\_e13.pdf](http://www.uneca.org/espd/statistics/docs/fasdev_2004_e13.pdf) (accessed 2009, September 8).
- Olken, F. and Gey, F. 2006. Social Science data librarianship – a university curriculum. *IASSIST Conference: Data in a World of Networked Knowledge*, Inter-university Consortium for Political and Social Research (ICPSR), University of Michigan, Ann Arbor, Michigan, May 23-26. [Online]. Available: [http://www.iassistdata.org/conferences/2006/presentations/A1\\_GuyOlken.pdf](http://www.iassistdata.org/conferences/2006/presentations/A1_GuyOlken.pdf) (accessed 2009, August 31).
- Onsembe, J.O., and Hie, J-M. 2004. The nature and magnitude of UNFPA's assistance to statistical development in Sub-Saharan Africa. In *Forum on African Statistical Development [FASDEV]*, Addis Ababa, Ethiopia, 12-13 May 2004. [Online]. Available: [http://www.uneca.org/espd/statistics/docs/fasdev\\_2004\\_e28.pdf](http://www.uneca.org/espd/statistics/docs/fasdev_2004_e28.pdf) (accessed 2009, August 31).
- Organisation for Economic Co-operation and Development. Development Co-operation Directorate (DCD-DAC). 2005. The Paris Declaration and AAA. [Online]. Available: [www.oecd.org/document/18/0,2340,en\\_2649\\_3236398\\_35401554\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/18/0,2340,en_2649_3236398_35401554_1_1_1_1,00.html) (accessed 2009, August 30).

## REFERENCES

- Organisation for Economic Cooperation and Development [OECD]. 2004. Science, Technology and Innovation for the 21st Century. Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level, 29-30 January 2004 - Final Communiqué: Annex 1: Declaration On Access To Research Data From Public Funding. [Online]. Available: [http://www.oecd.org/document/15/0,3343,en\\_2649\\_201185\\_25998799\\_1\\_1\\_1\\_1\\_00.html](http://www.oecd.org/document/15/0,3343,en_2649_201185_25998799_1_1_1_1_00.html) (accessed 2009, September 30).
- Organisation for Economic Cooperation and Development [OECD]. 2007. *OECD principles and guidelines for access to research data from public funding*. Paris: OECD. [Online]. Available: <http://www.oecd.org/dataoecd/9/61/38500813.pdf> (accessed 2009, August 31).
- Oslo Summer School in Comparative Social Science Studies [OSS]. 2009. Foundation, history, main objectives and target groups. [Online] Available at: <http://www.sv.uio.no/oss/oss01.html> (accessed 2009, August 30).
- Owino, A.Y. 2007. Institutional arrangements for statistical systems for socio-economic transformation in Africa. *Symposium on Science, Technology & Innovation: Key Drivers for Socio-Economic Transformation*, Kampala, Uganda, 13-15th September. [Online]. Available: <http://www.east-sym.org/documents/Owino.pdf> (accessed 2009, August 31).
- Page-Shipp, R. 2007. A Qualitative Study of the Management of Access to Research Data in South Africa: A prelude to a Policy Response to OECD Principles and Guidelines. A Position paper Commissioned by the Department of Science and Technology for a workshop on National Access to Research Data, 27-28 September.
- Partnership in Statistics for Development in the 21st Century (PARIS21). 2006. Review of support to statistical capacity building in Sub-Saharan Africa: from reporting to collaboration: synthesis report on the Light Reporting Exercise, Vol. I: final draft. [Online]. Available: <http://unstats.un.org/unsd/statcom/doc06/CapacityBuilding-Results-Paris21.pdf> (accessed 2009, August 31).

## REFERENCES

- Partnership in Statistics for Development in the 21st Century (PARIS21). 2008. National Strategies for the Development of Statistics: worldwide report on progress and emerging issues. Paris: PARIS21 Secretariat. [Online] Available: <http://www.paris21.org/documents/3323.pdf> (accessed 2009, August 31).
- Partnership in Statistics for Development in the 21st Century [PARIS21]. 2009a. About PARIS21. [Online]. Available: <http://www.paris21.org/pages/about-paris21/presentation/presentation> (accessed 2009, August 30).
- Partnership in Statistics for Development in the 21st Century [PARIS21]. 2009b. All events. [Online]. Available: [http://www.paris21.org/pages/events/all-events/list/?id\\_event=515](http://www.paris21.org/pages/events/all-events/list/?id_event=515) (accessed 2009, September 30).
- Porter, R.W. and Hicks, I. 2007. Knowledge utilization and the process of policy formation: towards a framework for Africa. Washington: Academy for Educational Development. [Online]. Available: [http://sara.aed.org/publications/cross\\_cutting/knowledge\\_utilization/html/utilization.htm](http://sara.aed.org/publications/cross_cutting/knowledge_utilization/html/utilization.htm) (accessed 2009, August 31).
- Ramachandran, V. 2006 ([vramachandran@worldbank.org](mailto:vramachandran@worldbank.org)). 2006. [Personal e-mail, 24 January] to Lynn Woolfrey ([lynn.woolfrey@uct.ac.za](mailto:lynn.woolfrey@uct.ac.za))
- Regional Reference Strategic Framework for Statistical Capacity Building in Africa [RRSF]. 2006. *African Statistical Journal* 2 (May):131-134. [Online] Available: [http://www.afdb.org/fileadmin/uploads/afdb/Documents/Publications/African.Statistical.Journal\\_Vol2\\_2.Articles\\_6.ReferenceRegionalStrategicFramework.pdf](http://www.afdb.org/fileadmin/uploads/afdb/Documents/Publications/African.Statistical.Journal_Vol2_2.Articles_6.ReferenceRegionalStrategicFramework.pdf) (accessed 2009, September 8).
- Regionalization of social sciences in Latin America, Asia and Africa. 1973. *International Social Science Journal* 25(4):557-560.
- Republic of South Africa. 2008. Intellectual property rights from publicly financed research and development Act [Act 51 of 2008]. [Online]. Available: <http://www.pmg.org.za/files/bills/081222a51-08.pdf> (accessed 2009, September 10).

## REFERENCES

- Republic of The Sudan and the Sudan People's Liberation Army. 2005. Comprehensive Peace Agreement between the Government of the Republic of Sudan and the Sudan People's Liberation Movement/Sudan People's Liberation Army. [Online]. Available: <http://www.aec-sudan.org/docs/cpa/cpa-en.pdf> (accessed 2009, August 30).
- Research Information Network. 2008. Stewardship of digital research data: a framework of principles and guidelines; responsibilities of research institutions and funders, data managers, learned societies and publishers. [Online]. Available: <http://www.rin.ac.uk/files/Research%20Data%20Principles%20and%20Guidelines%20of%20version%20-%20final.pdf> (accessed 2009, August 31).
- Review of support to statistical capacity building in Sub-Saharan Africa: from reporting to collaboration: synthesis report on the Light Reporting Exercise, Vol. I: final draft. 2006. [Online]. Available: <http://unstats.un.org/unsd/statcom/doc06/CapacityBuilding-Results-Paris21.pdf> (accessed 2009, August 31).
- Roper Center. 2009. Public opinion archives. [website] Available : <http://www.ropercenter.uconn.edu> (accessed 2009, August 30).
- Roper Public Opinion Research Center. 1963. *International Social Science Journal* 15(1):120-121.
- Rosen, B. and Elvers, E. 1997. Quality concept for official statistics. In Kotz, S, Read, C.B. and D.L.Banks, editors. *Encyclopedia of Statistical Sciences update volume 3*, 620-629. New York: John Wiley.
- Rothenbacher, F. 1994. Statistical sources for comparative European social research. *International Social Science Journal* 46(4):541-570.
- Rowe, J.S. 1974-75. Facilitating information access: interaction between system components (the data library and the traditional library). *ACM SIGSOC Bulletin* 6(2-3):32-38. [Online]. Available: <http://delivery.acm.org/10.1145/1110000/1102977/p32-rowe.pdf?key1=1102977&key2=6218219421&coll=GUIDE&dl=GUIDE&CFID=45979508&CFTOKEN=84657748> (accessed 2009, August 31).

## REFERENCES

- Rowe, J. S. 1976. National and international developments in providing data information services. In *Proceedings of the 4<sup>th</sup> annual ACM conference on User Services: Session 8c: data archives and data file services*, 97-99. [Online]. Available: <http://delivery.acm.org/10.1145/590000/584336/p97-rowe.pdf?key1=584336&key2=4231368421&coll=GUIDE&dl=GUIDE&CFID=45378518&CFTOKEN=51447139> (accessed 2009, August 31).
- Rowe, J.S. 1999. The decades of my life. *IASSIST Quarterly* 23(1):4-10.
- Rozsa, G. 1981. European cooperation in social science information and documentation: a process of maturation. *International Social Science Journal* 33(3):559-565.
- Rwandan national data center. 2009. *Balancing Act News Update* 443 February. [Online]. Available: [http://www.balancingact-africa.com/news/back/balancing-act\\_443.html](http://www.balancingact-africa.com/news/back/balancing-act_443.html) (accessed 2009, August 31).
- Scheuch, E.K. 1990. From a data archive to an infrastructure for the social sciences. *International Social Science Journal* 42(1):93 -111.
- Scheuch, E.K. 2003. History and visions in the development of data services for the social sciences. *International Social Science Journal* 55(177):385-399.
- School of Development Studies. University of Kwazulu-Natal. [2009?]. *KwaZulu-Natal Income Dynamics Study (KIDS) [Poverty and Inequality]*. [Online]. Available: <http://sds.ukzn.ac.za/default.php?7,12,9,4,0> (accessed 2009, September 30).
- Schouten, B. and Cigrang, M. 2003. Remote access systems for statistical analysis of microdata. *Statistics and computing* 13:381-389.
- Schurer, K. 2008. Social Science Implementation and management of a national initiative. *First African Digital Curation Conference*, Pretoria, South Africa 12-13 February.
- Seekings, J. 2002. *The uneven development of quantitative social science in South Africa*. Cape Town: Centre for Social Science Research. (CSSR Working Paper No. 6).

## REFERENCES

- Selematsela, D. 2009. Personal e-mail to Lynn Woolfrey regarding NeDCC database of digital curation institutions in Africa and data management courses in Southern Africa. (17 April 2009).
- Shibanda, G. 2006. Management of government information in Kenya. WSIS Follow-up Conference on Access to Information and Knowledge for Development, UN Conference Centre, Addis Ababa, Ethiopia, 27-30 March. [Online]. Available: <http://www.uneca.org/disd/events/2006/wsis-library/presentations/Management%20of%20Government%20Information%20in%20Kenya%20-%20George%20Shibanda%20-%20EN.pdf> (accessed 2009, August 31).
- Simonpietri, A. and Williams, T. 2005. Relevance of the National Strategy for the Development of Statistics (NSDS) approach to statistical development in Africa. *African Statistical Journal* 1:12-17.
- Söderberg, S.-J. 2005. MONA – Microdata On-line Access at Statistics Sweden. Joint UNECE/Eurostat work session on statistical data confidentiality, Geneva, Switzerland, 9-11 November, 2005. [Online]. Available: [www.unece.org/stats/documents/ece/ces/ge.46/2005/wp.3.e.pdf](http://www.unece.org/stats/documents/ece/ces/ge.46/2005/wp.3.e.pdf) (accessed 2009, August 31).
- South Africa. Department of Science and Technology [DST]. 2009. [website]. Centres of excellence. [Online]. Available: <http://www.dst.gov.za/centres-of-excellence> (accessed 2009, August 30).
- South Africa. National Archives and Record Service. [2009?]. About the National Archives and Record Service of South Africa. [web site] Available: [http://www.national.archives.gov.za/aboutnasa\\_content.html](http://www.national.archives.gov.za/aboutnasa_content.html) (accessed 2009, 5 September 2009).
- South Africa. Parliament. 1999. Statistics Act, 1999. [Online]. Available: [http://www.statssa.gov.za/about\\_statssa/statistics\\_act.asp](http://www.statssa.gov.za/about_statssa/statistics_act.asp) (accessed 13 July 2009).

## REFERENCES

- South African Data Archive. 1998. Proceedings of a workshop, "Data sharing for research capacity development", Pretoria, 9 July 1998. Pretoria: Human Sciences Research Council.
- South African Data Archive [SADA]. 2008. Introduction. [Online]. Available: <http://www.nrf.ac.za/sada/introduction.html> (accessed 2009, August 30).
- South African Data Archive [SADA]. 2009. Data catalogue. [Online]. Available: <http://www.nrf.ac.za/sada/ahlist.asp> (accessed 2009, August 30).
- Southern Africa Labour and Development Research Unit [SALDRU]. 2009. About us. [Online]. Available: <http://www.saldru.uct.ac.za> (accessed 2009, August 30).
- Southwood, R. 2008. Tomorrow's internet users today – African universities can play catch-up with online content. 2008. *Balancing Act News Update* 391. [Online]. Available: [http://www.balancingact-africa.com/news/back/balancing-act\\_391.html](http://www.balancingact-africa.com/news/back/balancing-act_391.html) (accessed 2009, August 31).
- Southwood, R. 2009. Fibre backbone cables spreading across the continent should lead to cheaper traffic between countries. *Balancing Act News Update* 459. [Online]. Available: [http://www.balancingact-africa.com/news/back/balancing-act\\_459.html](http://www.balancingact-africa.com/news/back/balancing-act_459.html) (accessed 2009, August 31).
- SPSS. 2009. About SPSS Inc.: Corporate history. [Online]. Available: <http://www.spss.com/corpinfo/history.htm> (accessed September 26).
- Stamatiou, W. 1988. International cooperation in the social sciences: the case of the Vienna Centre. *International Social Science Journal* 40(4): 597-603.
- Statistics Canada. 2000. *Policy on informing users of data quality*. Ottawa: Statistics Canada. [Online]. Available: [http://www.statcan.gc.ca/about-apercu/policy-politique/info\\_user-usager-eng.htm](http://www.statcan.gc.ca/about-apercu/policy-politique/info_user-usager-eng.htm) (accessed 2009, August 31).

## REFERENCES

- Statistics Canada. 2003. *Quality guidelines*. 2003. Ottawa: Statistics Canada. Available: <http://www.statcan.gc.ca/pub/12-539-x/12-539-x2003001-eng.pdf> (accessed 2009, August 31).
- Statistics South Africa [StatsSA]. 2008. South African statistical quality assessment framework. [Online]. Available: [http://www.statssa.gov.za/inside\\_statssa/standardisation/Statistics\\_SA\\_Statistical\\_Quality\\_Assessment\\_Framework.pdf](http://www.statssa.gov.za/inside_statssa/standardisation/Statistics_SA_Statistical_Quality_Assessment_Framework.pdf). (accessed 2009, August 31).
- Taamouti, M. 2009a. [Personal interview with Lynn Woolfrey, 17 August]. Durban. (Unpublished).
- Taamouti, M. ([mtaamouti@statistic.gov.za](mailto:mtaamouti@statistic.gov.za)). 2009b. [Personal e-mail, 7 September] to Lynn Woolfrey ([lynn.woolfrey@uct.ac.za](mailto:lynn.woolfrey@uct.ac.za)).
- Tanenbaum, E. and M. Taylor. 1991. Developing social science data archives. *International Social Science Journal* 43(1):225-234.
- Task Force on Higher Education and Society. 2000. *Higher education in developing countries: peril and promise*. 2000. Washington: World Bank. [Online]. Available: <http://www.tfhe.net/report/downloads/report/whole.pdf> (accessed 2009, September 20).
- Taylor, M. 1994. Ethical considerations in European cross-national research. *International Social Science Journal* 46(4):523-532.
- Tulya-Muhika, S. 1990. Teaching statistics for future government statistical services in Africa. In International Conference on Teaching Statistics (ICOTS) 3, Dunedin, New Zealand, 19-24 August 1990, 552-561. [Online]. Available: <http://www.stat.auckland.ac.nz/~iase/publications/18/BOOK2/C5-10.pdf> (accessed 2009, August 31).
- UK Data Archive [UKDA]. Good Practice in Data Documentation. 2002. Colchester: UK Data Archive. [Online]. Available: <http://www.esds.ac.uk/news/goodPractice.pdf> (accessed 2009, August 31).

## REFERENCES

- UK Data Archive [UKDA]. 2007: Research and development. [website]. Colchester. UKDA. [Online]. Available: <http://www.data-archive.ac.uk/ukda40/about/research.asp> (accessed 2009, August 31).
- United Kingdom. Office of Fair Trading. 2006. *The commercial use of public information*. London: Office of Fair Trading. [Online]. Available: [http://www.oft.gov.uk/shared\\_oftr/reports/consumer\\_protection/oft861.pdf](http://www.oft.gov.uk/shared_oftr/reports/consumer_protection/oft861.pdf) (accessed 2009, August 31).
- United Kingdom. Office of National Statistics [ONS]. 2009. Publication hub. [Online]. Available: <http://www.statistics.gov.uk> (accessed 2009, August 30).
- United Nations Department for Economic and Social Development. 2009. Financing for Development: latest developments [Online]. Available: <http://www.un.org/esa/ffd/index.htm> (accessed 2009, September 30).
- United Nations Economic and Social Council. 1997. Reform of the intergovernmental machinery of the Economic Commission for Africa: Resolution 1997/5. In 34th Plenary Meeting, 18 July 1997. [Online]. Available: <http://www.un.org/documents/ecosoc/res/1997/eres1997-5.htm> (accessed 2009, August 31).
- United Nations Economic and Social Council. Economic Commission for Africa. 1999. Developing National Information and Communication Infrastructure (NICI) policies, plans and strategies: the why and how. In First Meeting of the Committee on Development Information (CODI), Addis Ababa, Ethiopia. 28 June - 2 July 1999. [Online]. Available: <http://www.uneca.org/aisi/codipap2.htm> (accessed 2009, August 31).
- United Nations Economic and Social Council. Economic Commission for Africa. 2001. Assessment of the implementation of the Addis Ababa Plan of Action (AAPA) for Statistical Development in Africa in the 1990s. Presented at the Second meeting of the Committee on Development Information (CODI), Addis Ababa, Ethiopia, 4-7 September 2001. Document no. ECA/DISD/CODI.2/11. [Online]. Available: <http://www.uneca.org/codi/Documents/PDF/doc11en.pdf> (accessed 2009, August 30).

## REFERENCES

- United Nations Economic and Social Council. Economic Commission for Africa. 2007. National Information and Communication Infrastructure (NICI) e-Strategies: best practices and lessons learnt. Addis Ababa: Economic Commission for Africa. [Online]. Available: <http://www.uneca.org/aisi/nici/Docu> (accessed 2009, August 30).
- United Nations Economic and Social Council. Economic Commission for Africa. 2007. What is AISI? [Online]. Available: <http://www.uneca.org/aisi> (accessed 2009, September 30).
- United Nations Economic and Social Council. Economic Commission for Africa. 2008. The coordination of statistical capacity building efforts in Africa: an African perspective. In First joint annual meeting of the AU Conference of Ministers of Economics and Finance and ECA Conference of African Ministers of Finance, Planning and Economic Development, Addis Ababa, 26-29 March 2008. Document E/ECA/COE/27/13AU/CAMEF/EXP/13(III). [Online]. Available: <http://www.uneca.org/cfm/2008/docs/CoordinationofStatisticalCapacity.pdf>
- United Nations Economic and Social Council. Economic Commission for Africa. African Centre for Statistics [ACS]. 2009a. Events. [Online]. Available: <http://www.uneca.org/STATISTICS/symposium-rwanda.htm> (accessed 2009, September 30).
- United Nations Economic and Social Council. Economic Commission for Africa. African Centre for Statistics [ACS]. 2009b. News and information: 2008 African Statistics Day. [Online]. Available: <http://www.uneca.org/statistics> (accessed 2009, September 30).
- United Nations Economic and Social Council. Economic Commission for Africa. 2009a. Introduction: what is CODI? [Online]. Available: [http://www.uneca.org/codist/codi\\_tor.htm](http://www.uneca.org/codist/codi_tor.htm) (accessed 2009, September 30).
- United Nations Economic and Social Council. Economic Commission for Africa. 2009b. Overview of the ECA. [Online]. Available: <http://www.uneca.org/aisi> (accessed 2009, September 30).
- United Nations Economic and Social Council. Economic Commission for Africa. Forum on African Statistical Development [FASDEV]. 2004. [Online]. Available: <http://www.uneca.org/fasdev/fasdevl.asp> (accessed 2009, September 30).

## REFERENCES

- United Nations Economic and Social Council. Economic Commission for Africa. Forum on African Statistical Development [FASDEV]. 2006. FASDEV Newsletter II. (Friday 10 February). [Online]. Available: [http://www.uneca.org/fasdev/Documents/fasdevll\\_newsletter\\_closing.pdf](http://www.uneca.org/fasdev/Documents/fasdevll_newsletter_closing.pdf) (accessed 2009, September 10).
- United Nations Economic and Social Council. Economic Commission for Africa. Sub-Committee on Information and Communications Technology: ICT and Governance. 2008. e-Strategies: national, sectoral and regional ICT policies, plans and strategies. In Third meeting of the Committee on Development Information (CODI), Addis Ababa, 10-17 May 2003. [Online] Available: [http://www.uneca.org/eca\\_resources/publications/disd/e-strategies.pdf](http://www.uneca.org/eca_resources/publications/disd/e-strategies.pdf) (accessed 2009, August 31).
- United Nations Economic and Social Council. Economic Commission for Africa. Sustainable Development Division. [2005?]. Project proposal; Making Africa's power sector sustainable. [Online]. Available: [www.uneca.org/eca\\_programmes/sdd/documents/project\\_Making\\_Africa\\_power\\_sector\\_sustainable.htm](http://www.uneca.org/eca_programmes/sdd/documents/project_Making_Africa_power_sector_sustainable.htm) (accessed 2009, August 31).
- United Nations Educational, Scientific and Cultural Organization [UNESCO]. 2005. UNESCO science report. Paris: UNESCO. [Online]. Available: [http://www.unesco.org/science/psd/publications/science\\_report2005.pdf](http://www.unesco.org/science/psd/publications/science_report2005.pdf) (accessed 2009, August 31).
- United Nations Educational, Scientific and Cultural Organization [UNESCO]. 2009. Information for All Programme (IFAP). [Online]. Available: [http://portal.unesco.org/ci/en/ev.php-URL\\_ID=1627&URL\\_DO=DO\\_TOPIC&URL\\_SECTION=201.html](http://portal.unesco.org/ci/en/ev.php-URL_ID=1627&URL_DO=DO_TOPIC&URL_SECTION=201.html) (accessed 2009, September 20).
- United Nations Population Information Network [POPIN]. Regional Institute for Population Studies [RIPS]. [2003?]. Ghana: summary information. [Online]. Available: <http://www.un.org/popin/popdiv/tsspop/handlist/africa/ghana.htm> (accessed 2009, August 30).

## REFERENCES

- United Nations. Statistical Commission. 1994. Fundamental principles of official statistics and related sets of principles. [website]. Geneva: United Nations Economic Commission for Europe. [Online]. Available: <http://unstats.un.org/unsd/methods/statorg/FP-English.htm> (accessed 2009, August 31).
- United Nations Statistical Commission. [2009?]. Population and housing censuses. [Online]. Available: <http://unstats.un.org/unsd/demographic/sources/cwp2010/default.htm> (accessed 2009, September 30).
- University of Pretoria and University of Botswana. 2009. Conference fees and payments. In *African Digital Scholarship and Curation Conference*, Pretoria, 12-14 May 2009. [Online]. Available: <http://www.ais.up.ac.za/digi/fees.htm> [accessed 2009, September 9].
- US Census Bureau. Methodology and Standards Council. 2006a. Definition of data quality: Census Bureau principle. United States Census Bureau. [Online]. Available: [http://www.census.gov/quality/P01-0\\_v1.3\\_Definition\\_of\\_Quality.pdf](http://www.census.gov/quality/P01-0_v1.3_Definition_of_Quality.pdf) (accessed 2009, August 30).
- US Census Bureau. Methodology and Standards Council. 2006b. Development of concepts and methods. [Online]. Available: [http://www.census.gov/quality/P02-0\\_v1.2\\_Develop\\_Concepts.pdf](http://www.census.gov/quality/P02-0_v1.2_Develop_Concepts.pdf) (accessed 2009, August 30).
- US Census Bureau. Methodology and Standards Council. 2006c. Dissemination of Statistical Information to the Public: US Census Bureau Principle. United States Census Bureau. [Online]. Available: [http://www.census.gov/quality/P02-0\\_v1.2\\_Dissemination.pdf](http://www.census.gov/quality/P02-0_v1.2_Dissemination.pdf) (accessed 2009, August 30).
- Vahed, A. 2005. Towards a 21st Century technological platform for improving data quality in Statistics South Africa. *14th Conference of Commonwealth Statisticians*, Cape Town, 5-9 September, 2005. [Online]. Available: [http://www.statssa.gov.za/commonwealth/presentations/Paper\\_A\\_Vahed.pdf](http://www.statssa.gov.za/commonwealth/presentations/Paper_A_Vahed.pdf) (accessed 2009, August 31).

## REFERENCES

- van Rosendaal, C. J. 1984. European information policy situation. *Aslib Proceedings* 36(1):15–23.
- Verma, V. 2002. Comparability in international survey statistics [keynote address]. *International Conference on Improving Surveys*, Copenhagen, 25-28 August, 2002. [Online]. Available: [http://www.nessie-essex.co.uk/roundtable\\_1VermaFULL.pdf](http://www.nessie-essex.co.uk/roundtable_1VermaFULL.pdf) (accessed 2009, August 31).
- Weis, P. 2004. Borders in cyberspace: conflicting government information policies and their economic impacts. In Esanu, J.M. and Uhler, P.F. Editors. *Open access and the public domain in digital data and information for science*, 69-73. Washington: National Academies Press. Available: [http://books.nap.edu/openbook.php?record\\_id=11030&page=69](http://books.nap.edu/openbook.php?record_id=11030&page=69) [2009, August 31].
- Welch, M. 2007. Personal interview with Lynn Woolfrey (18 October).
- Williams, M. 2006. Broadband for Africa: Policy for promoting the development of backbone networks. Washington: Infodev. [Online]. Available: <http://www.infodev.org/en/Publication.526.html> (accessed 2009, August 31).
- Wilson, F. 1996. South Africa: poverty under duress. In *Poverty: a global review: handbook on international poverty research*, edited by E. Oyen, S. M. Miller and S.A. Samad, 227-247. Oslo: Scandinavian University Press.
- Wittenberg, M. 2006. Research note: errors in the October Household Survey 1994 available from the South African Data Archive. *South African Journal of Economics* 74(4): 766-768.
- Woolfrey, L. 2007. The establishment of the African Association of Statistical Data Archivists. *IASSIST Quarterly* 31(2):24-26.
- World Summit on the Information Society. 2003. Civil Society Declaration: “Shaping Information Societies for human needs”. [Online]. Available: <http://www.itu.int/wsis/docs/geneva/civil-society-declaration.pdf> (accessed 2009, September 15).

## REFERENCES

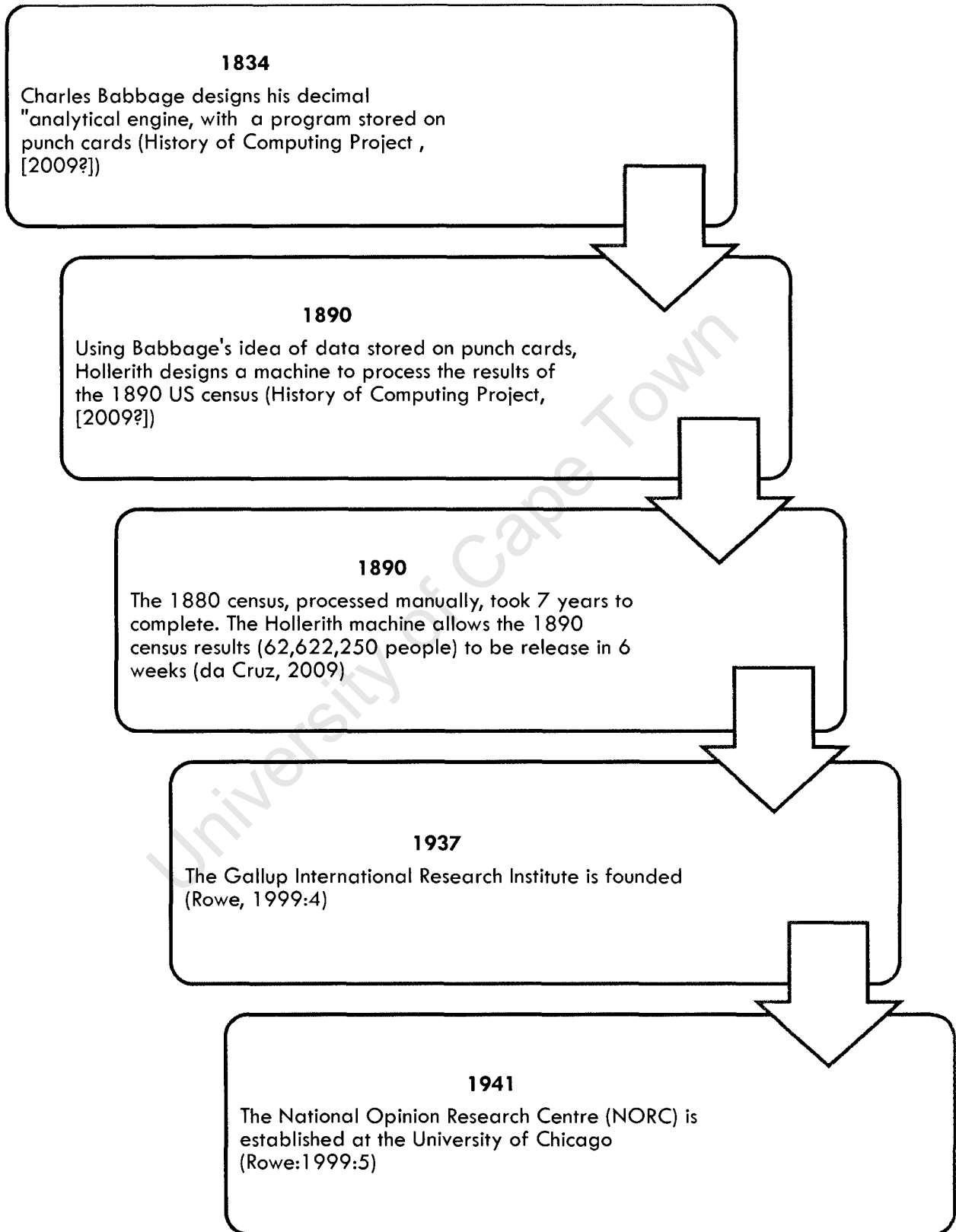
- World Values Survey [WVS]. 2009. Introduction to the World Values Survey. [Online].  
Available: <http://www.worldvaluessurvey.org> (accessed 2009, September 30).
- Yumba, D. 2002. Whither Africa in the Information Age? *Information Development* 18(4):237-243.

University of Cape Town

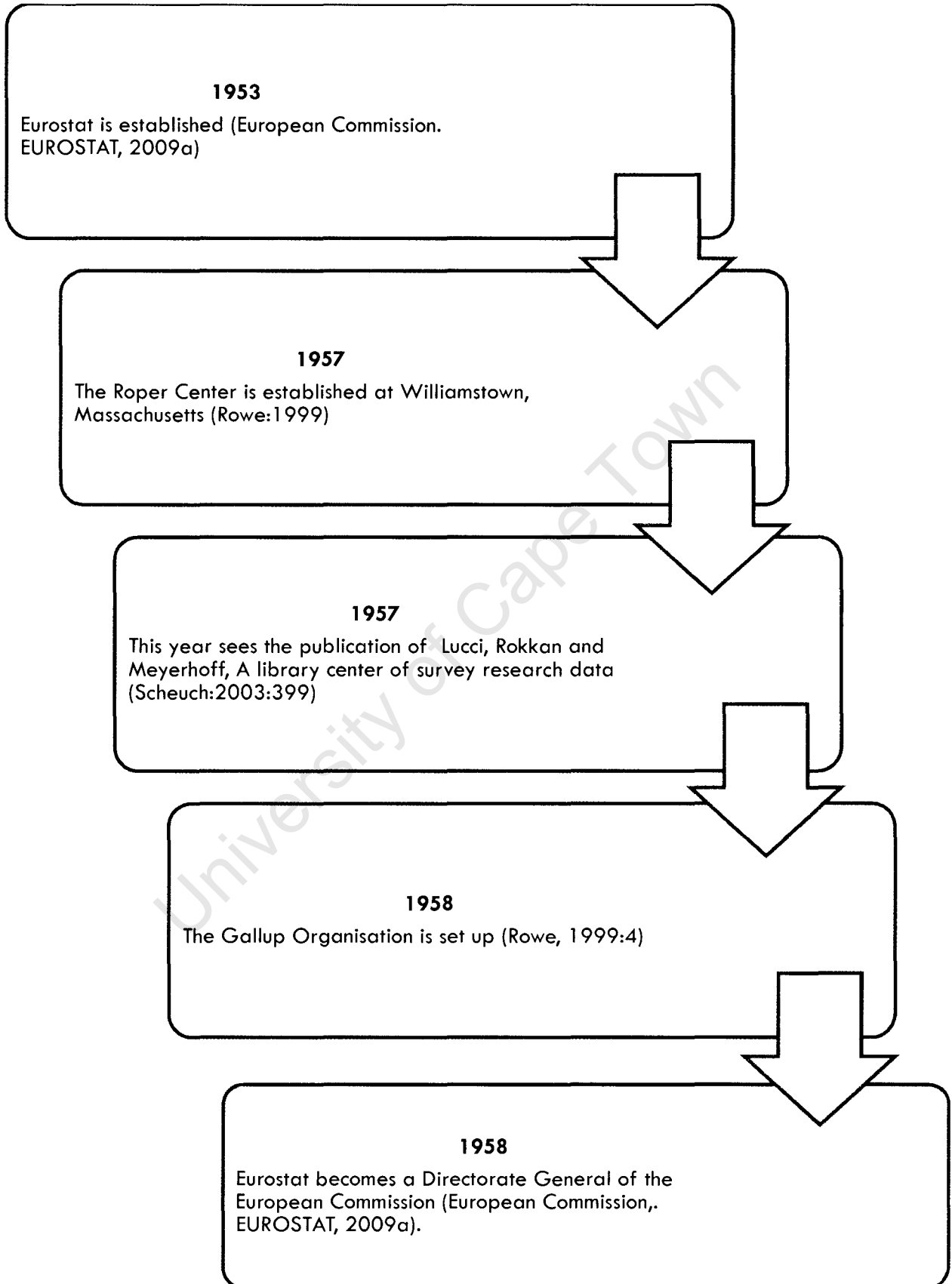
## APPENDICES

University of Cape Town

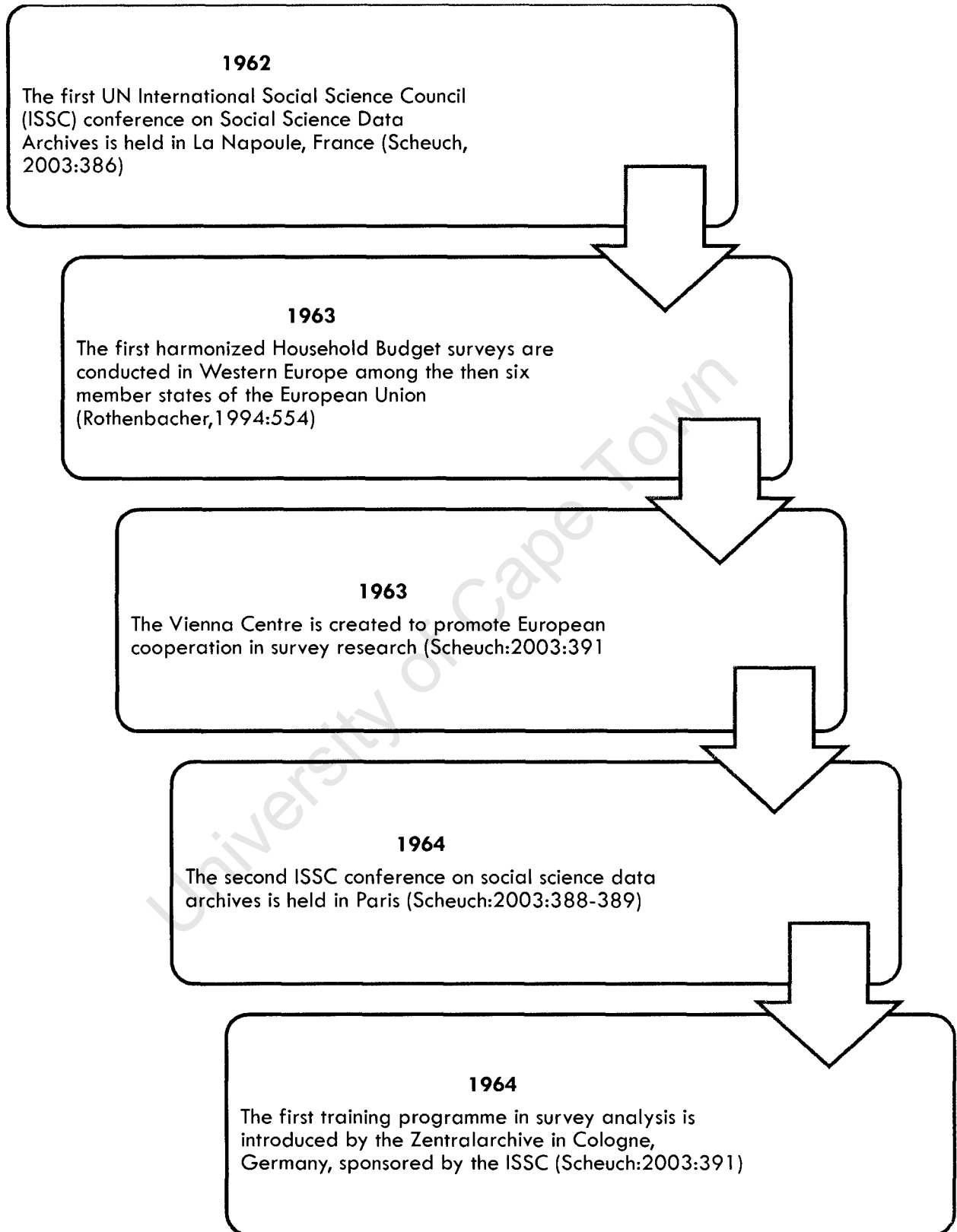
**Appendix A: Timeline Of Survey Data Curation Worldwide**



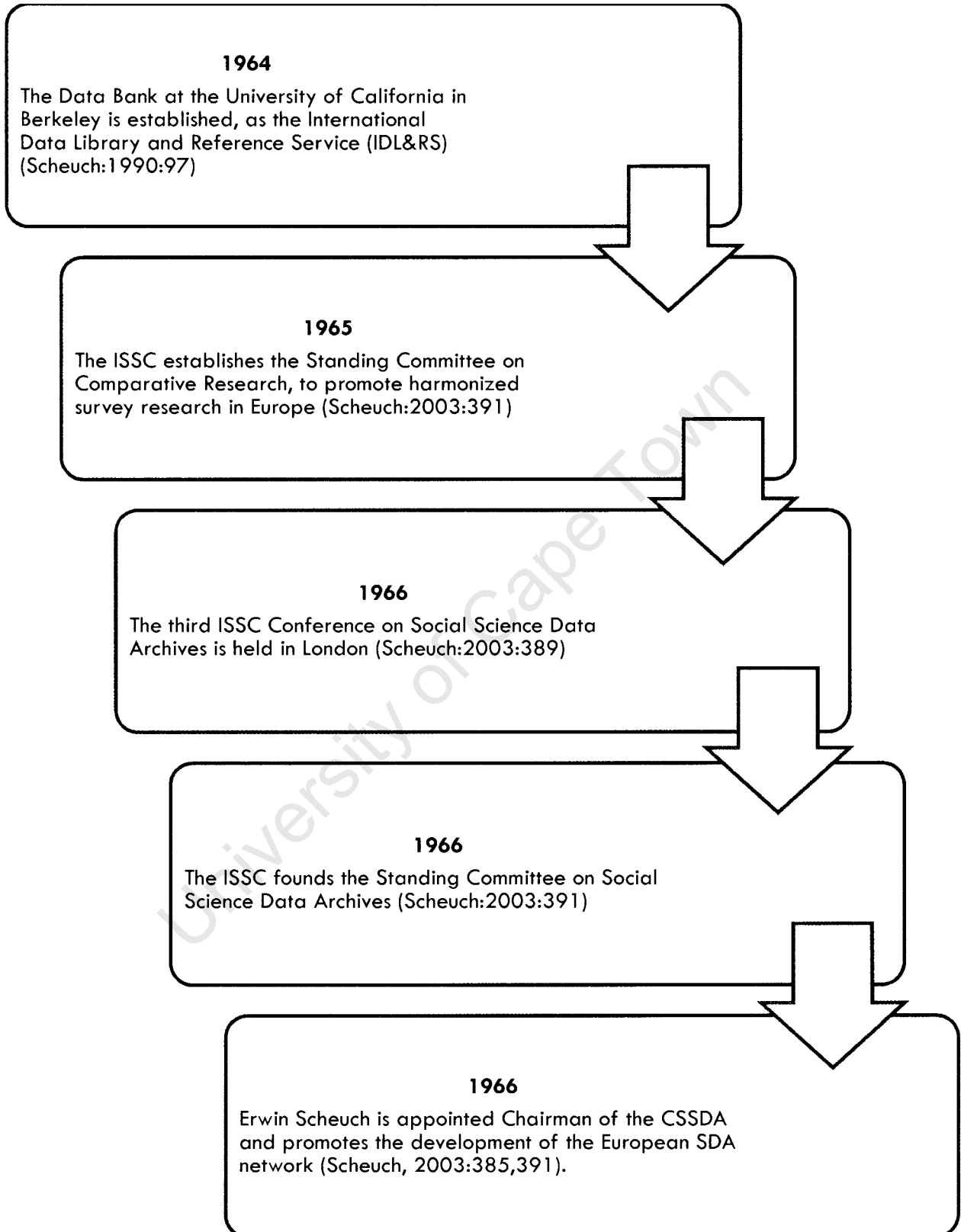
## DATA CURATION TIMELINE



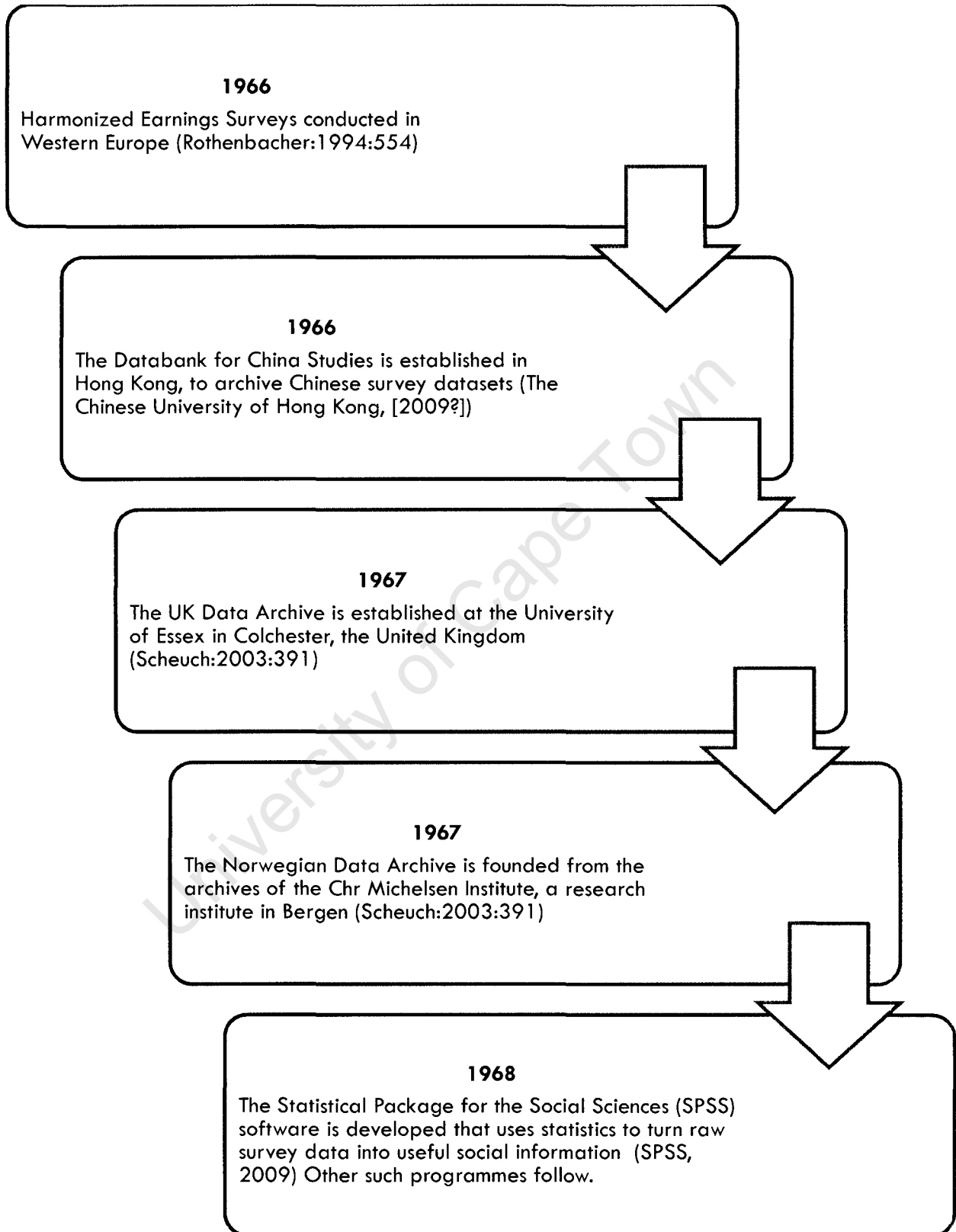
## DATA CURATION TIMELINE



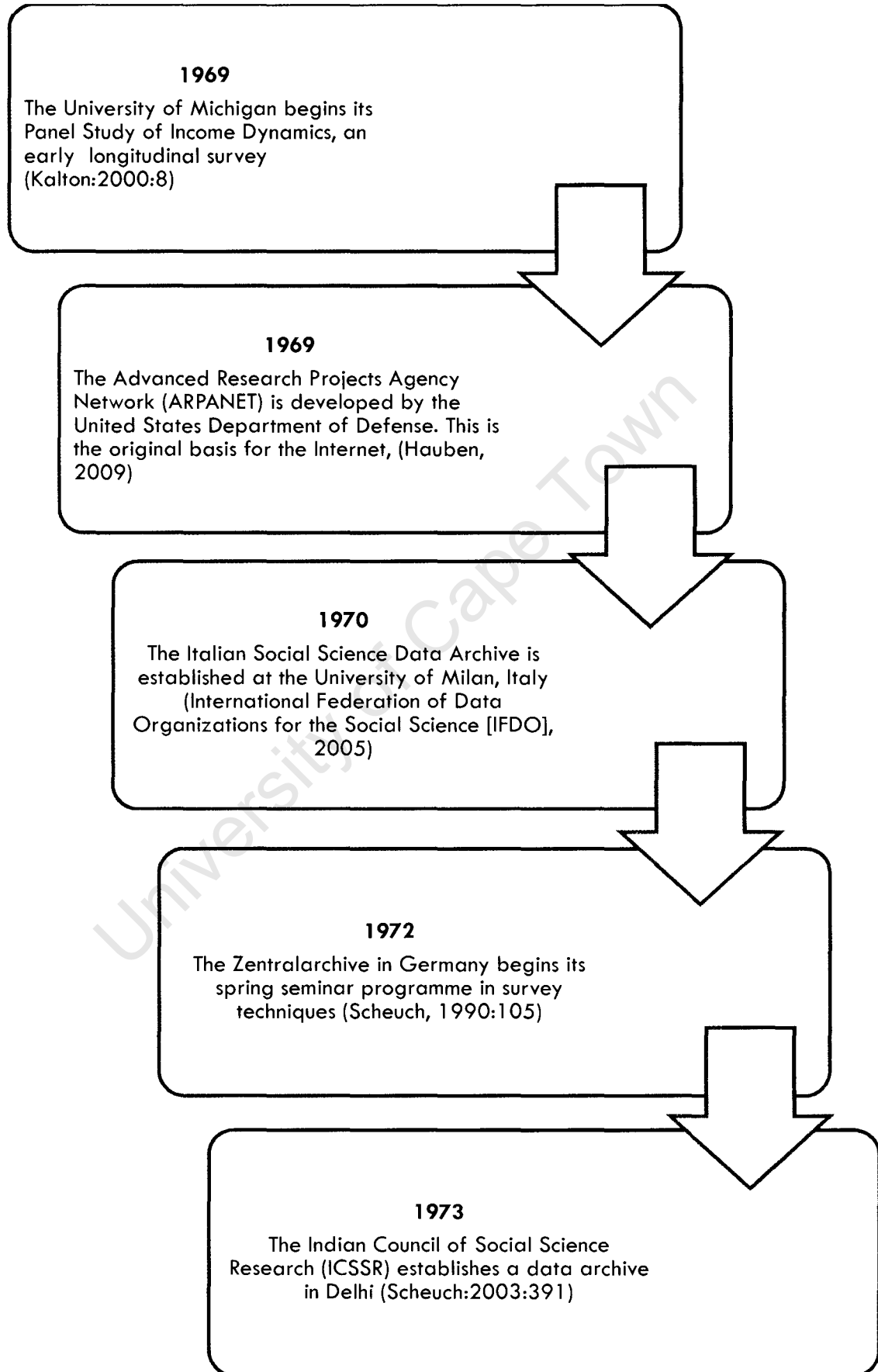
## DATA CURATION TIMELINE



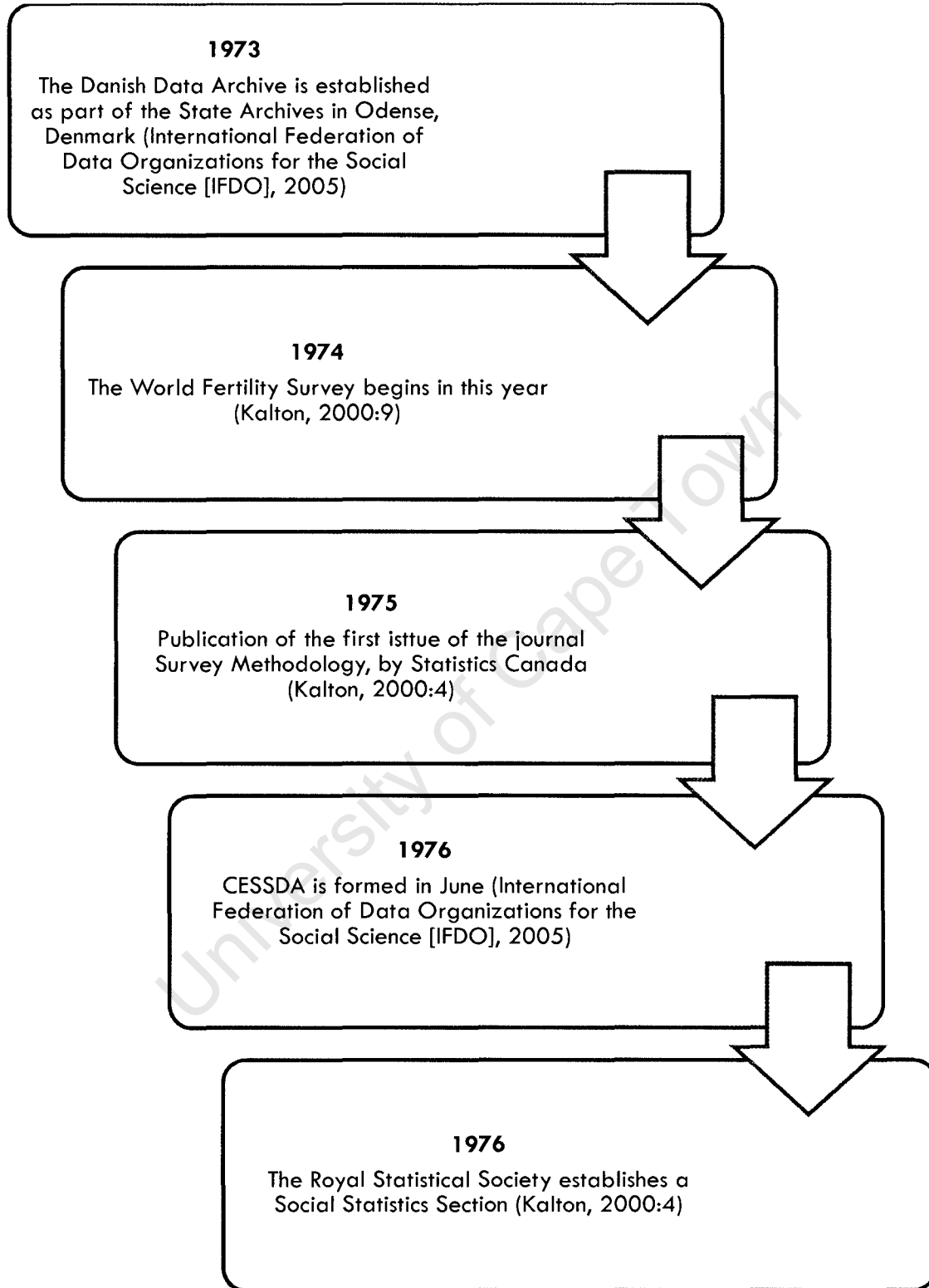
## DATA CURATION TIMELINE



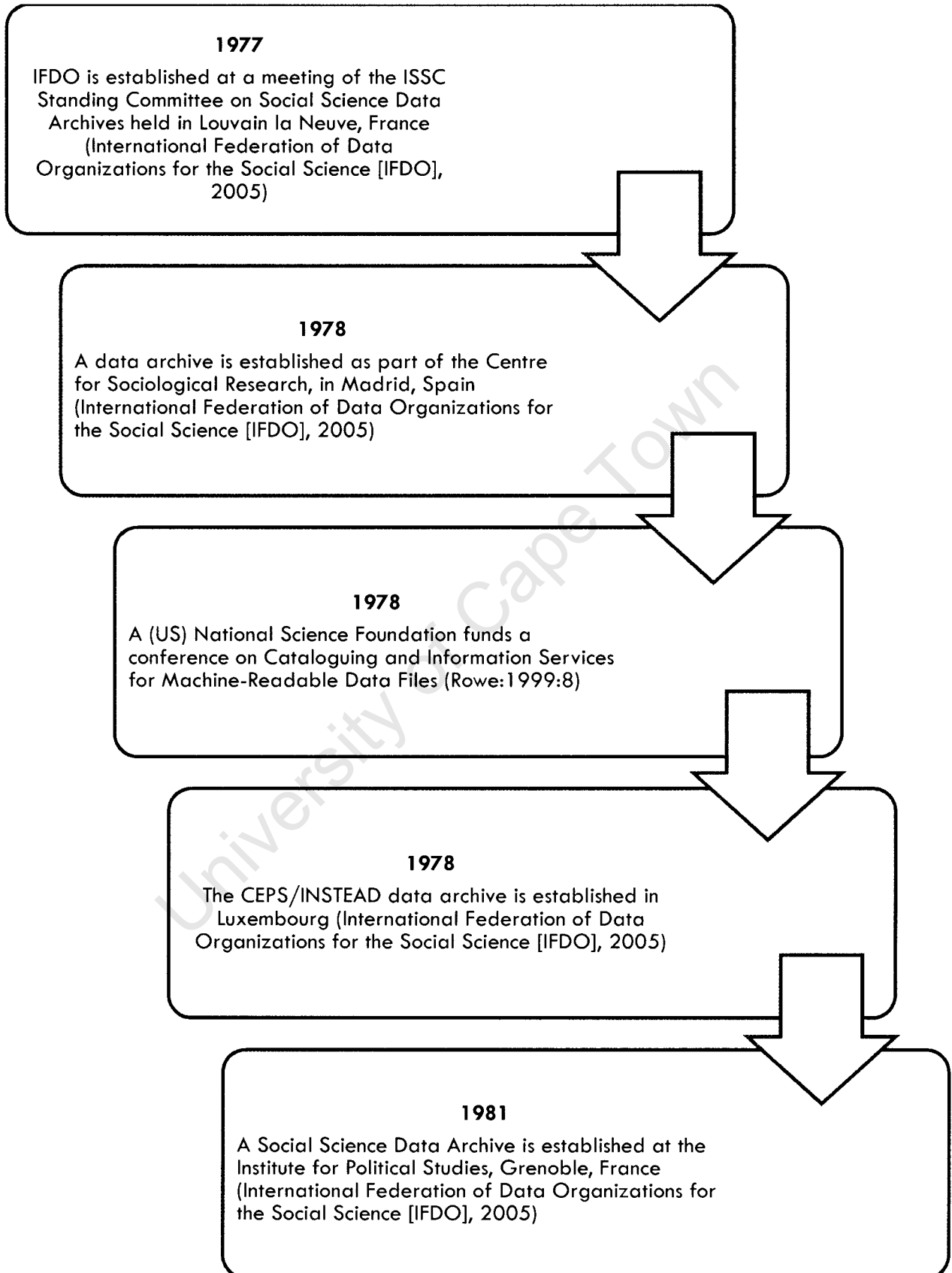
## DATA CURATION TIMELINE



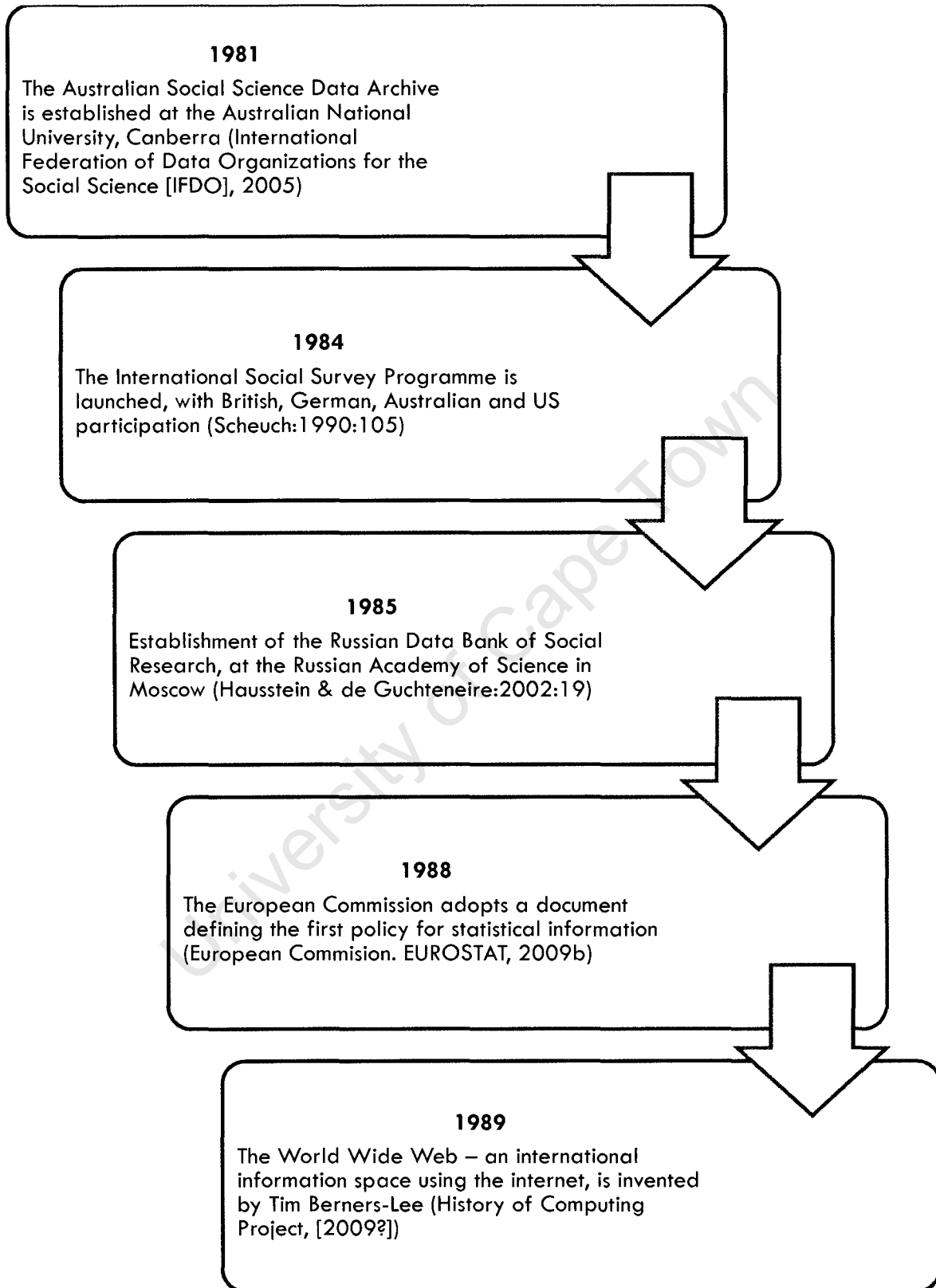
## DATA CURATION TIMELINE



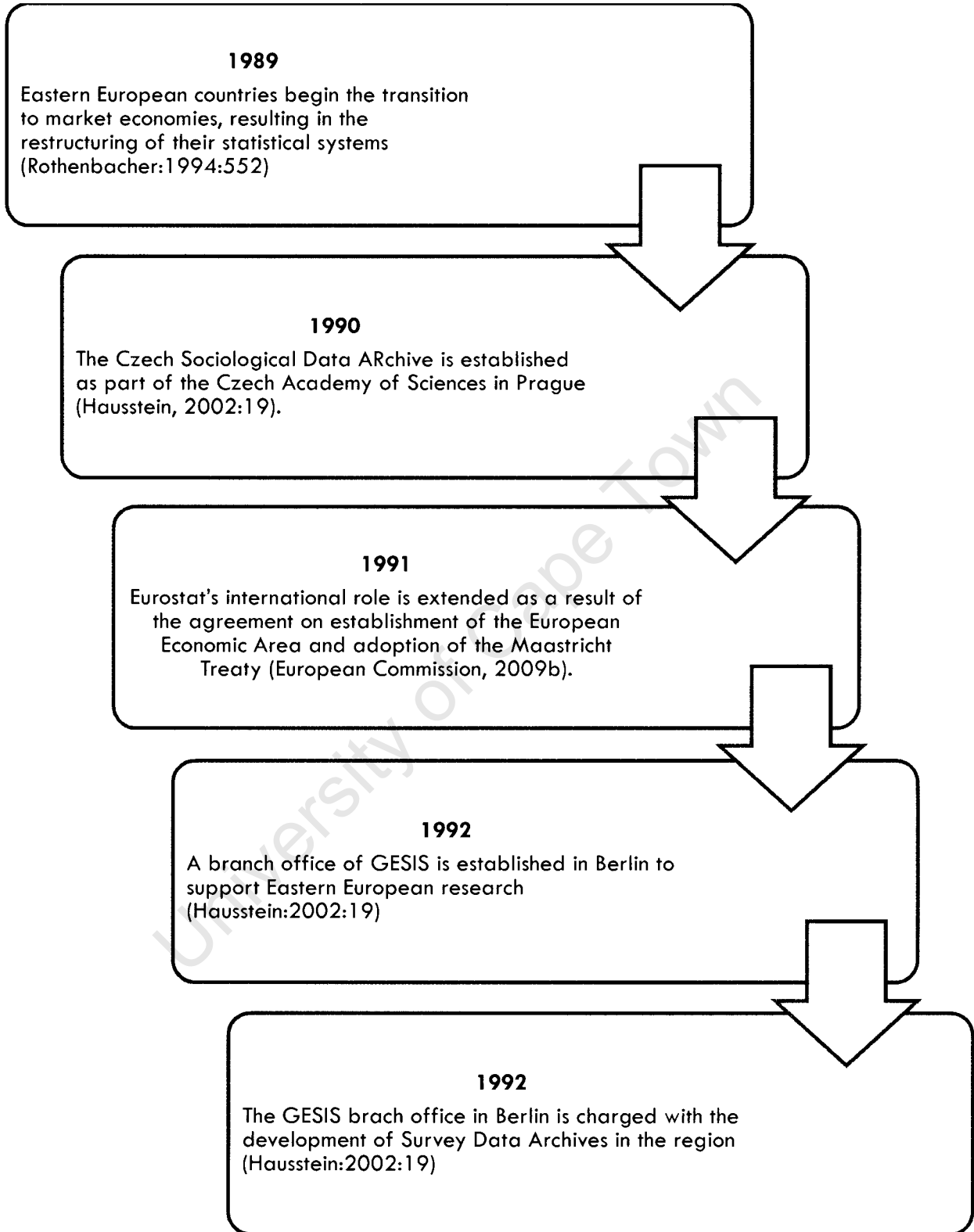
## DATA CURATION TIMELINE



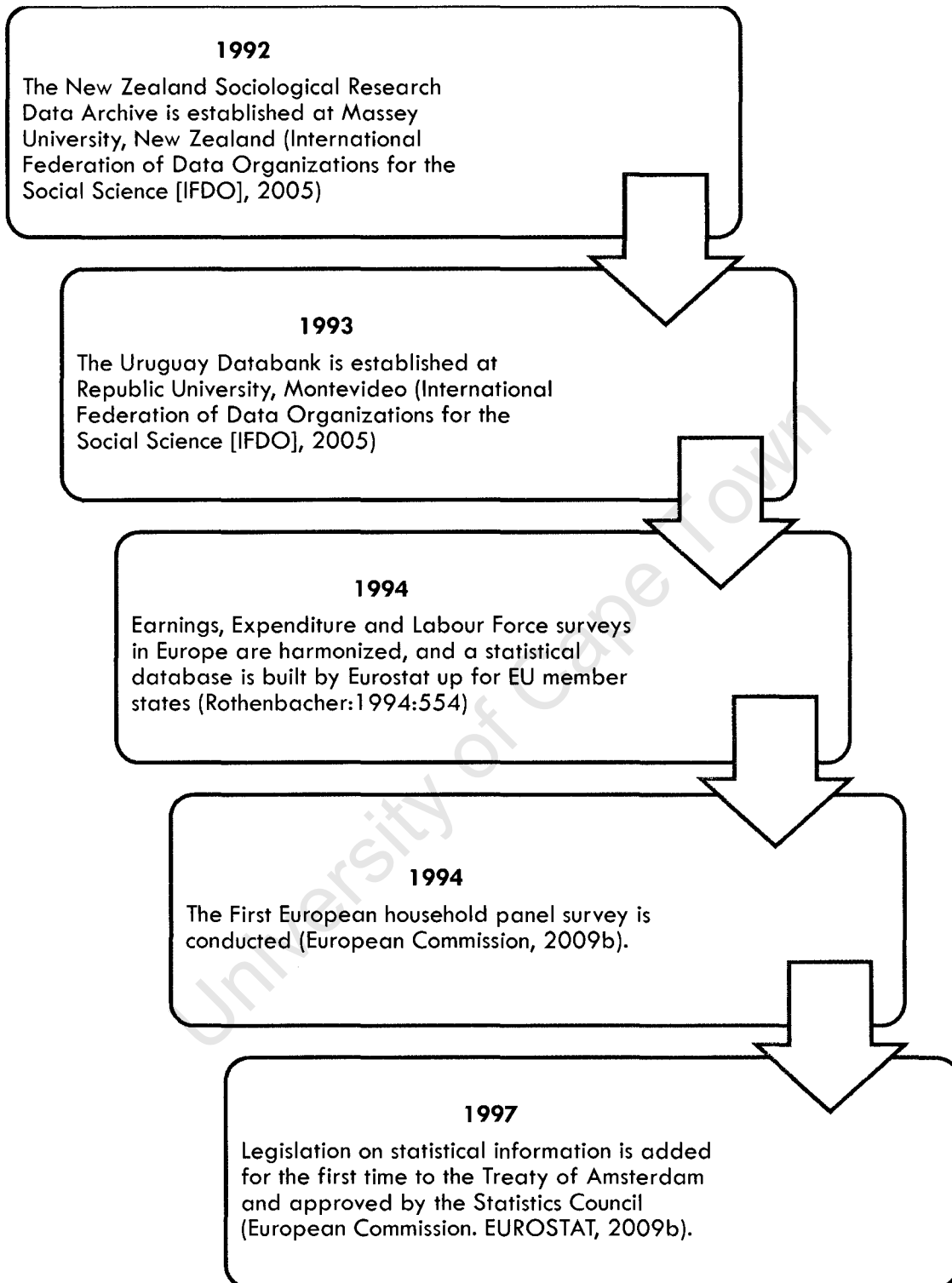
## DATA CURATION TIMELINE



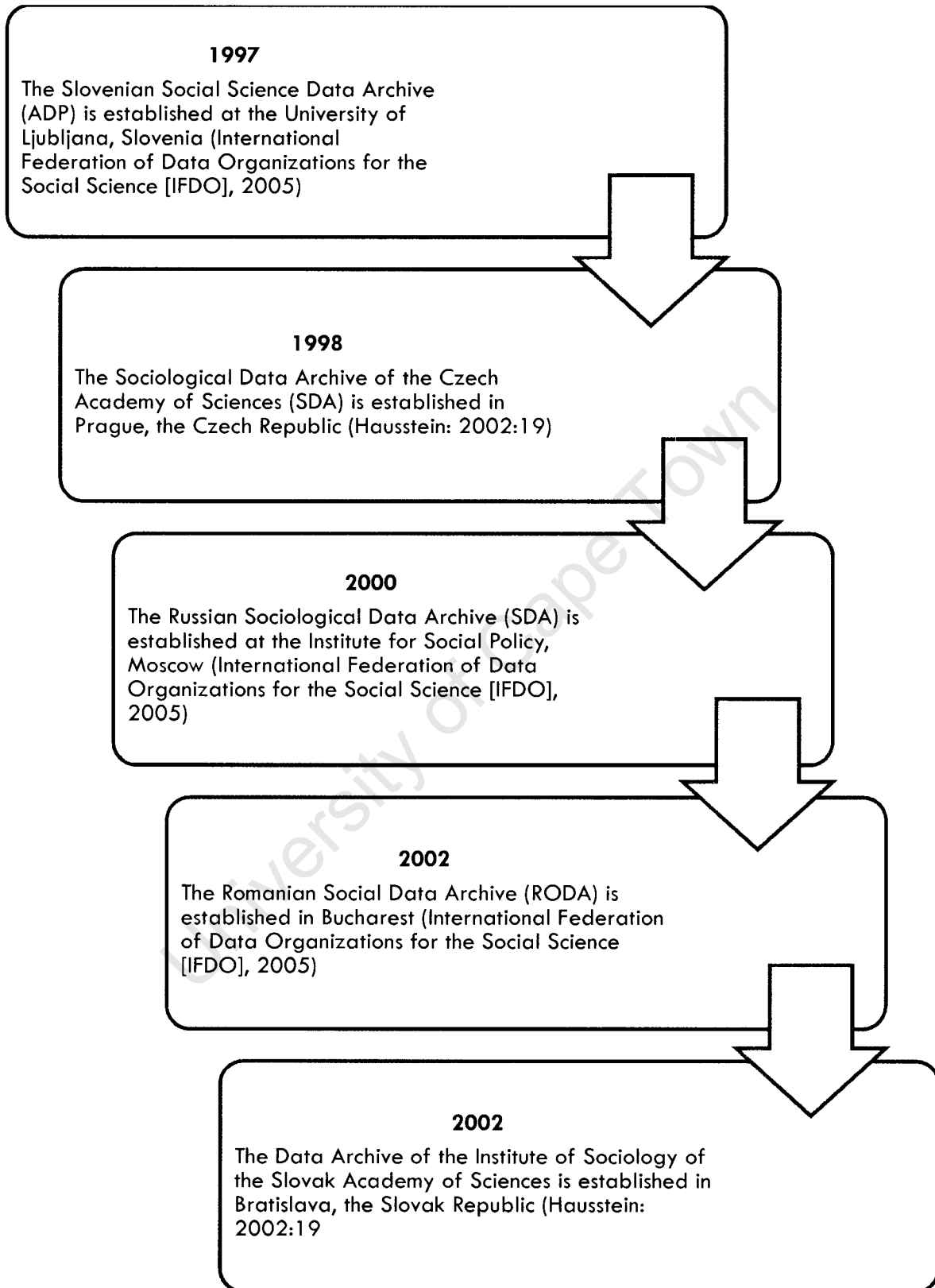
## DATA CURATION TIMELINE



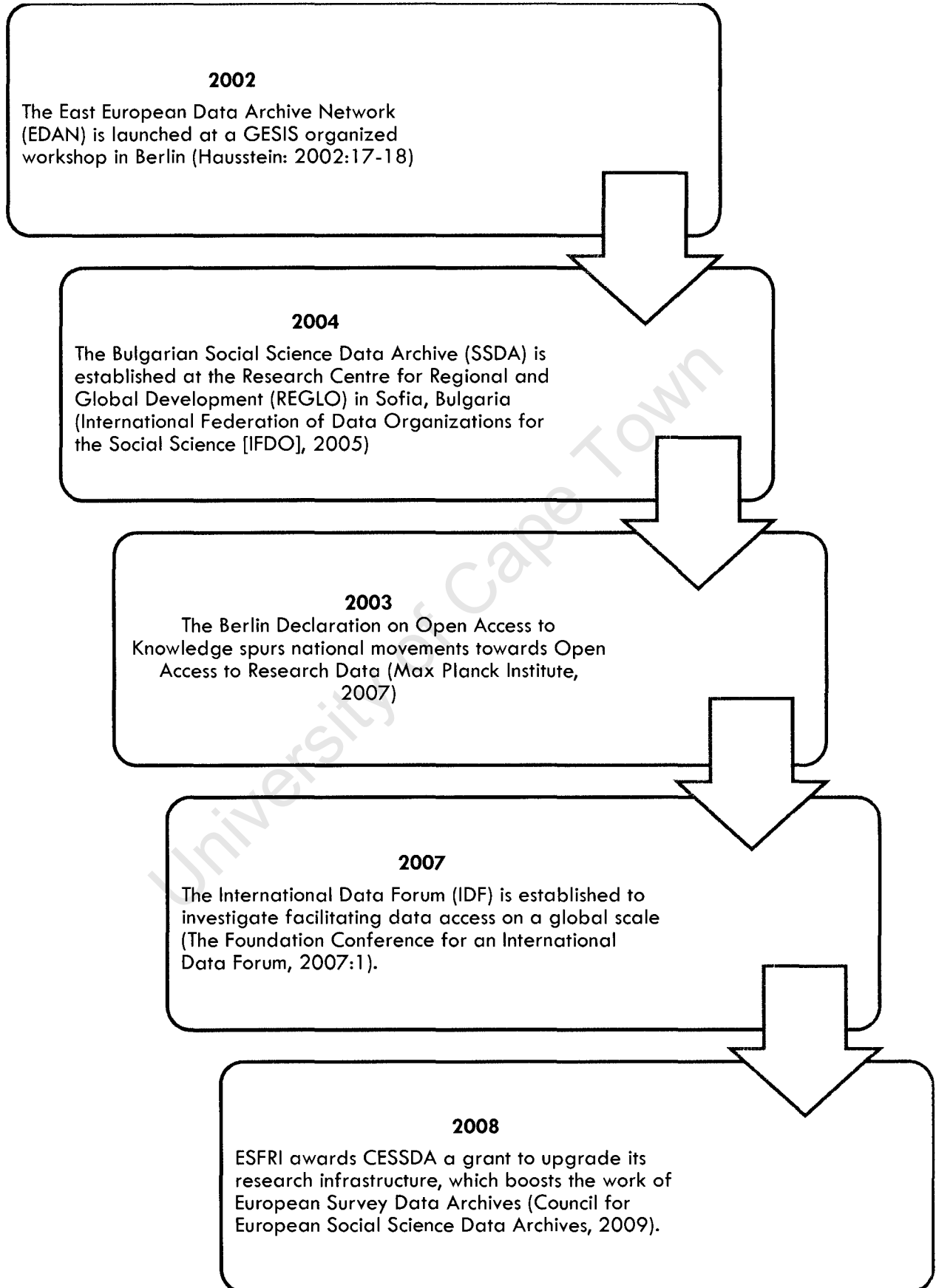
## DATA CURATION TIMELINE



## DATA CURATION TIMELINE



## DATA CURATION TIMELINE












SURVEY DATA ARCHIVES WORLDWIDE

Appendix B: Survey Data Archives Worldwide

COUNTRY	NAME	ABBREVIATION		INSTITUTION	URL	AFFIL.	EST.
Australia	Australian Social Science Data Archive	SSDA		Australian National University, Canberra	<a href="http://www.assda.anu.edu.au">http://www.assda.anu.edu.au</a>	IFDO	1981
Austria	Vienna Institute for Social Science Data and Methods	WISDOM		Vienna, Austria	<a href="http://wisdom.at">http://wisdom.at</a>	IFDO, CPSSA	1985
Belgium	Belgian Archives for the Social Sciences	BASS		Catholic University of Louvain-la-Neuve	<a href="http://logi.rspo.ucl.ac.be">http://logi.rspo.ucl.ac.be</a>	IFDO	1966
Brazil	Centre for the Study of Public Opinion	CESOP		Located at the Universidade Estadual de Campinas, São Paulo	<a href="http://www.cesop.unicamp.br">http://www.cesop.unicamp.br</a>		
Bulgaria	The Bulgarian Social Science Data Archive	BSSDA		Based in the Research Centre Regional and Global Development (REGLO), Sofia, Bulgaria	<a href="http://www.reglo-bg.org">http://www.reglo-bg.org</a>	IFDO, EDAN	2004
Canada	Statistics Canada	STATCAN		This is the government statistical agency, located in Ottawa, Canada	<a href="http://www.statcan.ca">http://www.statcan.ca</a>	IFDO	
Czech Republic	Sociological Data Archive	SDA		Czech Academy of Sciences, Prague	<a href="http://archiv.soc.cas.cz">http://archiv.soc.cas.cz</a>	IFDO, EDAN	1990
China	Databank for China Studies	DCS			<a href="http://www.usc.cuhk.edu.hk/databank">http://www.usc.cuhk.edu.hk/databank</a>		1966
Denmark	Danish Core Archive	DDA			<a href="http://www.ddc.dk">http://www.ddc.dk</a>	IFDO, CPSSA	1973






SURVEY DATA ARCHIVES WORLDWIDE

COUNTRY	NAME	ABBREVIATION		INSTITUTION	URL	AFFIL.	EST.
Estonia	Estonian Social Science Data Archive	ESSDA		Located in the Faculty of Social Science at Tartu University, Tartu	<a href="http://www.psych.ut.ee/esto">http://www.psych.ut.ee/esto</a>	CESSDA	1996
Finland	Finnish Social Science Data Archive	FSD		The Archive is a separate unit of the University of Tampere	<a href="http://www.fsd.uta.fi">http://www.fsd.uta.fi</a>	Estonia	1999
France	Social Science Data Archive (BDSP)	BDSP		The archive is located in the Institute for Political Studies in Grenoble (ICDSEI)	<a href="http://www-bdsp.upmf-grenoble.fr/cidsp">http://www-bdsp.upmf-grenoble.fr/cidsp</a>	FDG, CESSDA	198
Germany	Central Archive for Empirical Social Research	ZA		A unit of the Kölner Institute for Social Research (KIS) at the Univ. of Cologne	<a href="http://www.gesis.org/en/za">http://www.gesis.org/en/za</a>	IFDQ, CFSSDA, FDAN	1960
Greece	Greek Social Data Bank	GSDB		National Centre for Social Research (EKKE), Athens	<a href="http://www.ekke.gr/databank">http://www.ekke.gr/databank</a>	CESSDA	1998
Hungary	Social Research Informatics Centre	TARKI		Set up by the Major Research Organisations based in Budapest	<a href="http://www.tarki.hu">http://www.tarki.hu</a>	IFDQ, CFSSDA, ICPSR, ECPR	1985
India	National Social Science Documentation Centre	NASSDOC		A division of the Indian Council of Social Science Research, New Delhi	<a href="http://www.icssr.org">http://www.icssr.org</a>	IFDQ	1969
Ireland	Irish Social Science Data Archive	ISSDA		A unit of the Geary Institute for Social Sciences, University College Dublin	<a href="http://www.ucd.ie/issda">http://www.ucd.ie/issda</a>	CESSDA, IASJIST	1999
Israel	Israeli Social Sciences Data Centre	ISDC		Based at the Faculty of Social Sciences, Hebrew University, Jerusalem	<a href="http://isdc.huji.ac.il">http://isdc.huji.ac.il</a>	IFDQ	
Italy	Social Science Data Archive	ADPSS		Department of Sociology and Social Research, University of Milan	<a href="http://www.sociologia.unimi.it/sociodata">http://www.sociologia.unimi.it/sociodata</a>	IFDQ, CESSDA	1970






SURVEY DATA ARCHIVES WORLDWIDE

COUNTRY	NAME	ABBREVIATION	INSTITUTION	URL	AFFIL.	EST.
Jamaica	Derek Gordon Data Bank		 University of the West Indies, Mona, Jamaica, Sir Arthur Lewis Institute of Social and Economic Studies	<a href="http://www.uwi.tt/sslises/institute">http://www.uwi.tt/sslises/institute</a>		1999
Japan	Social Science Japan Data Archive	SSJDA	 Univ of Tokyo, Institute of Social Science, Information Centre for Social Science	<a href="http://ssjda.iss.u-tokyo.ac.jp">http://ssjda.iss.u-tokyo.ac.jp</a>		1998
Korea	Korean Social Science Data Centre	KSDC	 Seoul, South Korea	<a href="http://www.ksdc.re.kr">http://www.ksdc.re.kr</a>	ICPSR	1997
Latvia	Latvian Databank of Social Sciences	IS7DA	Academy of Sciences, Riga, Latvia			
Luxembourg	International Network for Studies in Technology, Environment, Alternatives, Development Luxembourg	CEPS/INSTEAD	A public non-profit organisation in Differdange, Luxembourg	<a href="http://www.ceps.lu">http://www.ceps.lu</a>		1978
Netherlands	Data Archiving and Networked Services	DANS	Since 2005 DANS (in The Hague) has been the umbrella data archive in the Netherlands	<a href="http://www.dans.knaw.nl">http://www.dans.knaw.nl</a>	IFDO, CE5SDA	2005
New Zealand	New Zealand Social Research Data Archive	NZSRDA	Located in the Faculty of Social Sciences, Massey University, New Zealand	<a href="http://www.massey.ac.nz/~nzsrda/nzsrda/archive">http://www.massey.ac.nz/~nzsrda/nzsrda/archive</a>	IFDO	1992
Norway	Norwegian Social Science Data Archive	NSD	Falls under the Research Council of Norway, with the head office at University of Bergen	<a href="http://www.nsd.uib.no">http://www.nsd.uib.no</a>	IFDO, CE5SDA	1971
Poland	Polish Social Data Archive	ADS	 Institute for Social Studies (ISS), Warsaw University, Warsaw	<a href="http://www.ads.org.pl">http://www.ads.org.pl</a>	EDAN	

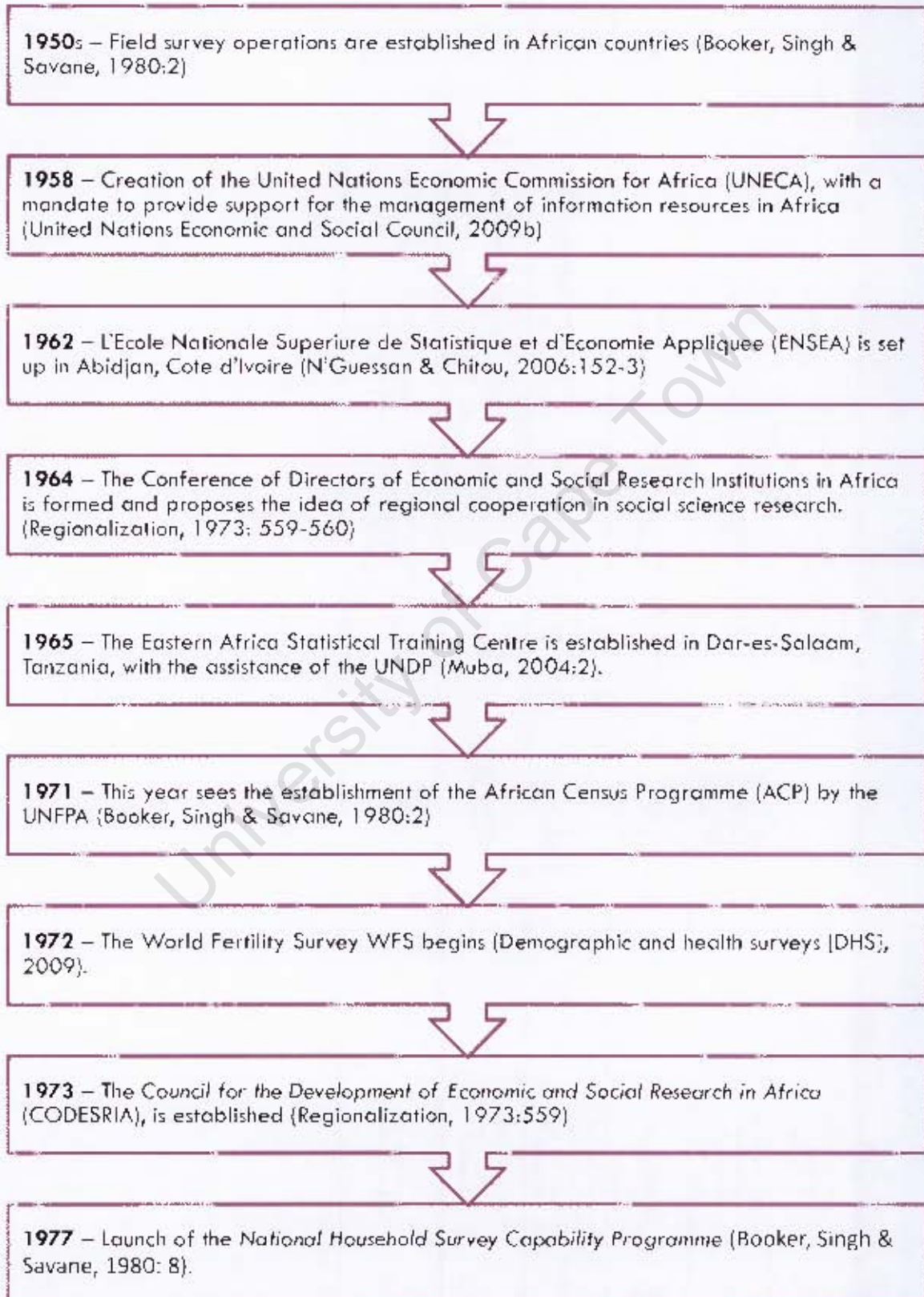
SURVEY DATA ARCHIVES WORLDWIDE

COUNTRY	NAME	ABBREVIATION		INSTITUTION	URL	AFFIL.	EST.
Russia	Data Bank of Social Research	DBSR		An academic non-profit research center within the Institute of Sociology of the Russian Academy of Science, Moscow	<a href="http://www.gesis.org/en/data_service/eastern_europe">http://www.gesis.org/en/data_service/eastern_europe</a>	EDAN	1985
	The Russian Sociological Data Archive	SDA		Part of the Independent Institute for Social Policy in Moscow	<a href="http://www.socialpolicy.ru">http://www.socialpolicy.ru</a>	EDAN	2000
Serbia	The Data Archive of the Centre for Political Studies and Public Opinion Research	CPJRM		Based in the Centre for Political Studies and Public Opinion Research, Belgrade, Serbia	<a href="http://www.cpijrm.org.yu/cpijrm.htm">http://www.cpijrm.org.yu/cpijrm.htm</a>	IFDC, EDAN	1963
Slovak Republic	Slovak Archive of Social Data	SASD		Department of Sociology, Faculty of Philosophy, Comenius Univ., Bratislava	<a href="http://sasd.konzum.sk">http://sasd.konzum.sk</a>	CESSDA, EDAN	2004
	Data Archive of the Institute for Sociology			The Institute of Sociology of the Slovak Academy of Sciences, Bratislava	<a href="http://www.sociologia.sav.sk/archiv">http://www.sociologia.sav.sk/archiv</a>	EDAN	1990 Insti
Slovenia	Slovenian Social Science Data Archive	ADP		Located in the Faculty of Social Sciences, University of Ljubljana	<a href="http://www.adp.fdv.uni-lj.si">http://www.adp.fdv.uni-lj.si</a>	EDAN	1997
South Africa	South African Data Archive	SADA			<a href="http://www.nrf.ac.za/sada/">http://www.nrf.ac.za/sada/</a>	ICPSR, FDO	1993
Spain	Centre for Sociological Research	CIS			<a href="http://www.cis.es">http://www.cis.es</a>	CESSDA	1977
Sweden	Swedish Social Science Data Archive	SSD		Part of the Social Science Faculty, University of Göteborg	<a href="http://www.ssd.gu.se">http://www.ssd.gu.se</a>	FDO, CESSDA	1980
Switzerland	Swiss Information and Data Archive Service	SIDOS		Part of the Swiss Academy of Humanities and Social Sciences	<a href="http://www.sidos.ch">http://www.sidos.ch</a>	FDO, CESSDA	1993

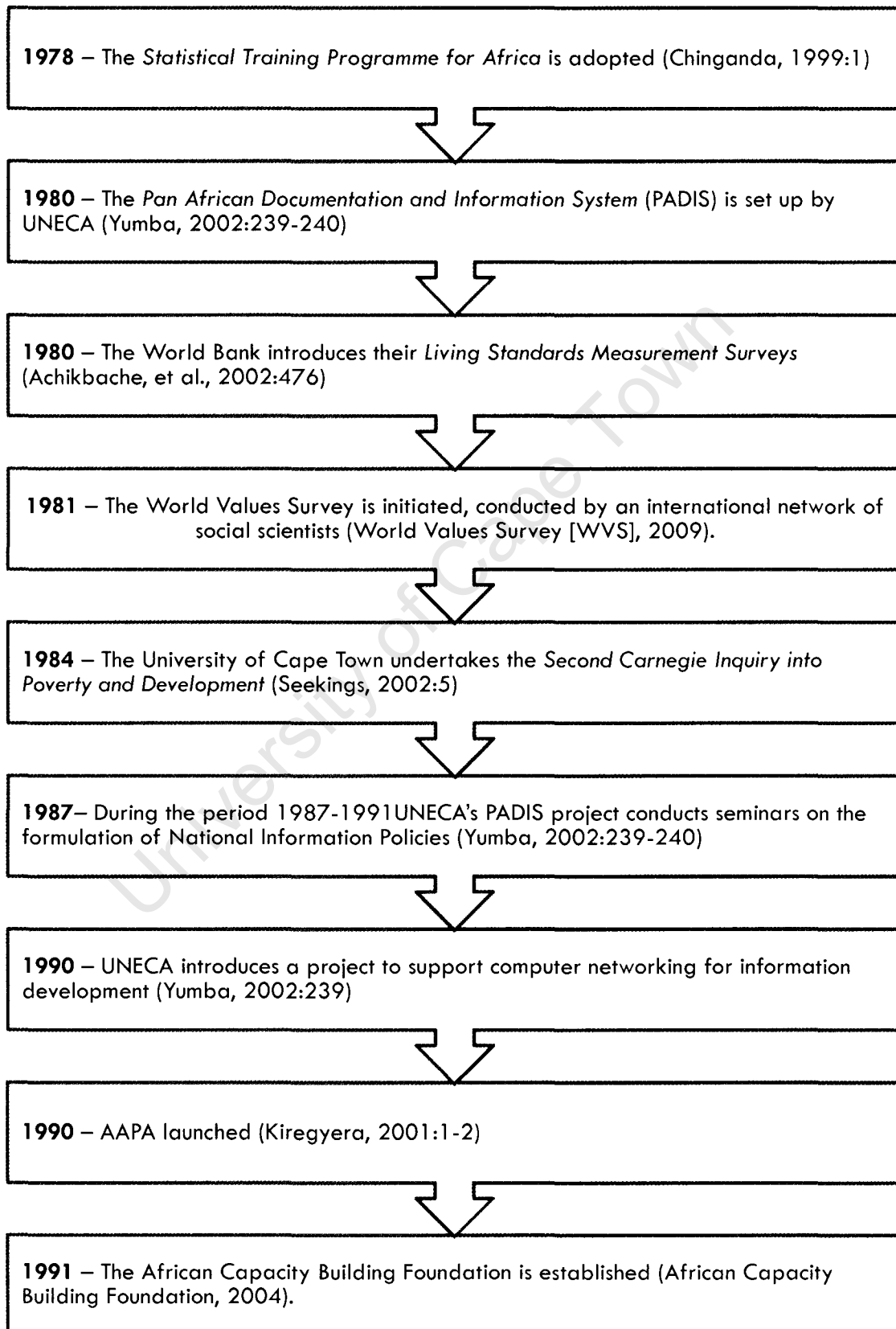
SURVEY DATA ARCHIVES WORLDWIDE

COUNTRY	NAME	ABBREVIATION	INSTITUTION	URL	AFFIL.	EST.
Taiwan	Center for Survey Research, Academia Sinica		A division of Academia Sinica, Taipei, Taiwan	<a href="http://www.sinica.edu.tw">http://www.sinica.edu.tw</a>	ICPSR Roper	1994
Ukraine	Databank of the International Institute of Sociology	KIIS	 Part of the International Institute of Sociology, Kiev	<a href="http://www.kiis.com.ua">http://www.kiis.com.ua</a>	EDAN	2001
United Kingdom	UK Data Archive	UK/ESRI	 The UK Data Archive is based at Essex University, Colchester, UK	<a href="http://www.data-archive.ac.uk">http://www.data-archive.ac.uk</a>	IFDO, CESSDA, ICPSR	1967
United States	Inter-university Consortium for Political and Social Research	ICPSR	 Located in the Institute for Social Research (ISR), University of Michigan	<a href="http://www.icpsr.umich.edu">http://www.icpsr.umich.edu</a>	IFDO	1962
	The Roper Center for Public Opinion Research		 University of Connecticut, Storrs	<a href="http://www.ropercenter.uconn.edu">http://www.ropercenter.uconn.edu</a>		1947
Uruguay	Data Bank – Banco de Datos		 Located in the Faculty of Social Sciences of the Republic University, Montevideo	<a href="http://www.rau.edu.uy/fcs/banco">http://www.rau.edu.uy/fcs/banco</a>	IFDO	1993

**Appendix C: African Data Curation Timeline**



## AFRICAN DATA CURATION TIMELINE



## AFRICAN DATA CURATION TIMELINE

**1992** – UNECA initiates African Statistics Day (United Nations Economic and Social Council, 2009b)

**1993** – AFRISTAT is established by the governments of West African States (Economic and Statistical Observatory of Sub-Saharan Africa [AFRISTAT]. 2009)

**1995** – *The African Regional Symposium on Telematics for Development* (United Nations Economic and Social Council 1999:2)

**1995** – *Towards Africa's Information Highway resolution tabled* (United Nations Economic and Social Council, 1999:2)

**1995** – The ADB undertakes a detailed assessment of the national statistical capacities of 32 African countries (Kiregyera, 2001:1-2).

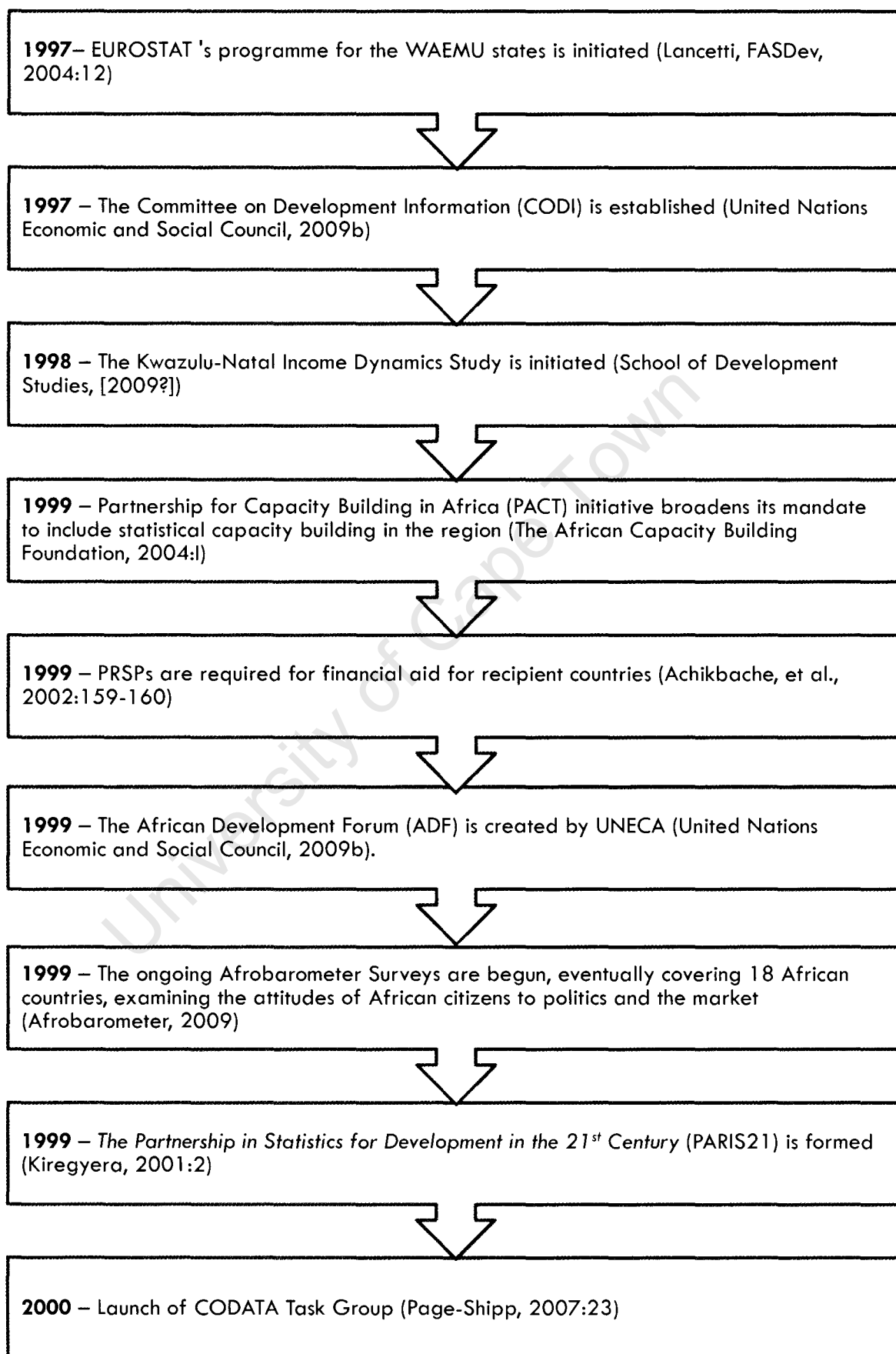
**1995** – The G7 Conference pledges support for African statistical development (Nassimbeni, 1998:158)

**1996** – The African Information Society Initiative (AISI) is launched (United Nations Economic and Social Council. Economic Commission for Africa. [2007?]b).

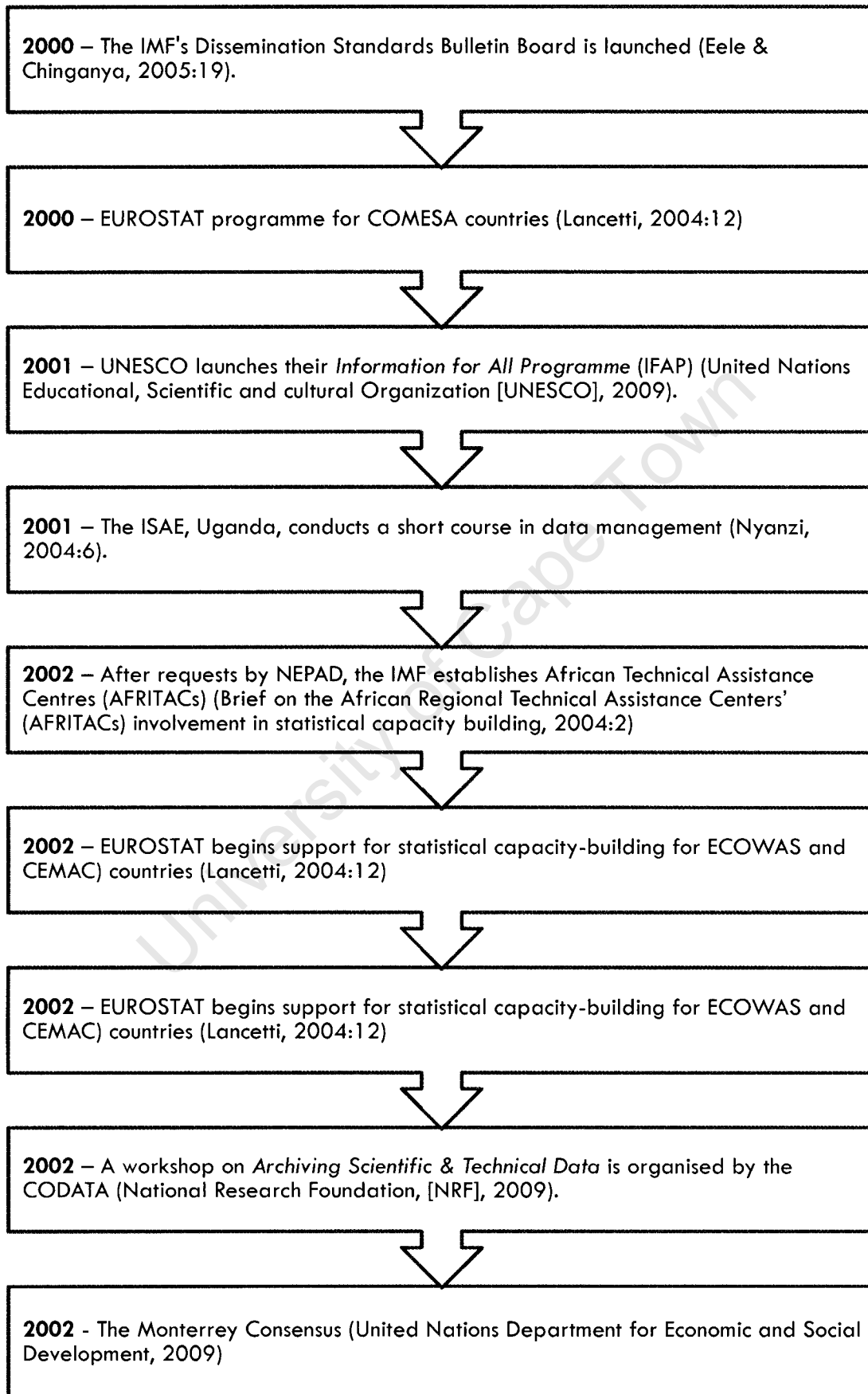
**1996** – EUROSTAT's MEDSTAT programme is launched (European Commission. EUROSTAT, 2009c)

**1997** – The IMF introduces the GDDS (Achikbache, et al., 2002:472)

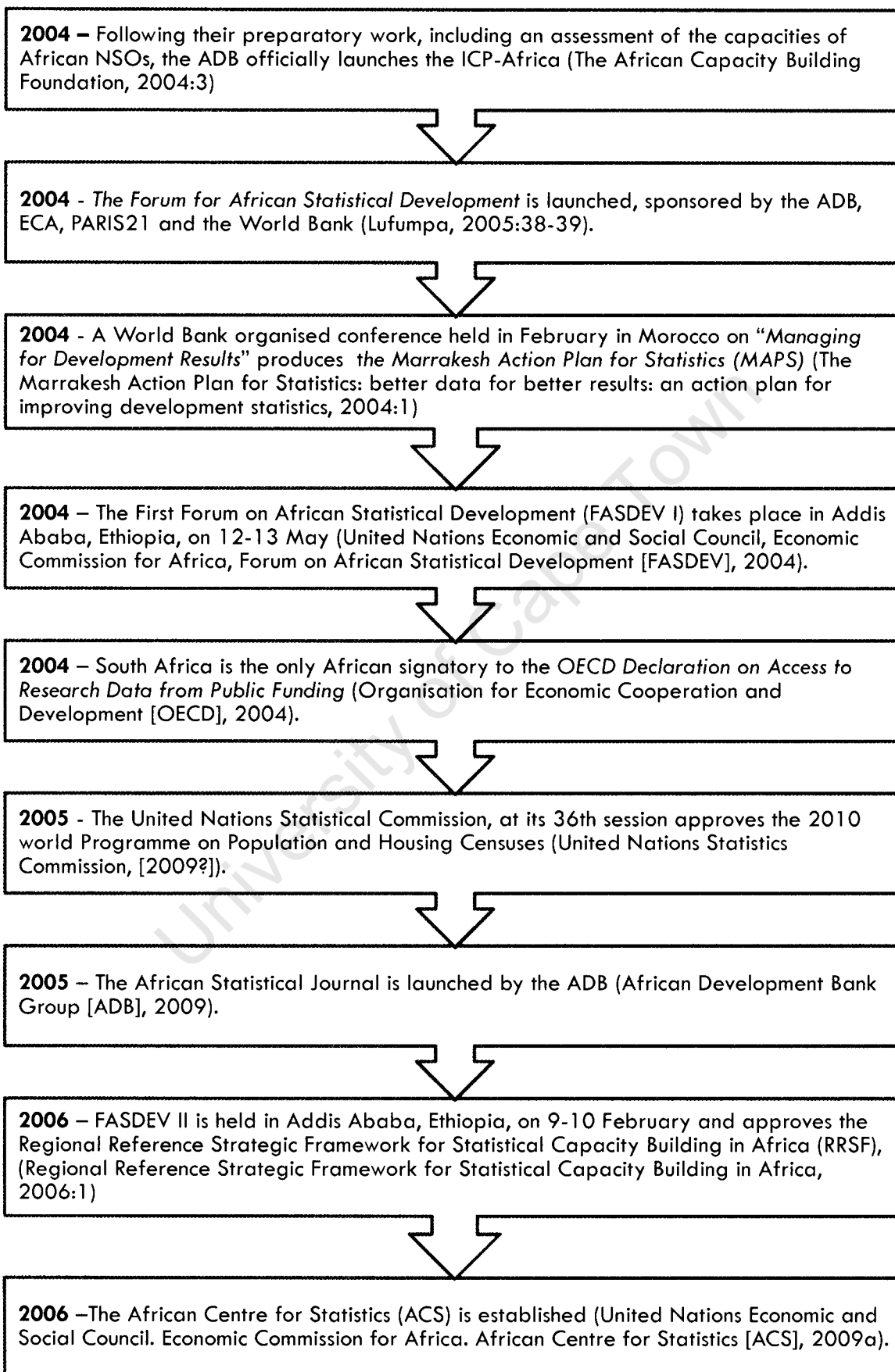
## AFRICAN DATA CURATION TIMELINE



## AFRICAN DATA CURATION TIMELINE



## AFRICAN DATA CURATION TIMELINE



## AFRICAN DATA CURATION TIMELINE

**2006** – The Inaugural African Symposium on Statistical Development (ASSD) is held in Cape Town, South Africa, 30 January – 2 February (Africa Symposium on Statistical Development, 2006a:4).

**2006** – The West African statistical support organisation, AFRISTAT, launches its AFRISTAT Strategic Plan of Activities (ASPA) 2006-2010 (Balepa, 2006:135-6)

**2006** – The Accelerated Data Programme (ADP) is launched to support the aims of the International Household Survey Network (IHSN) in promoting data preservation and dissemination in African (Dupriez, 2006:8-11)

**2007** – The Second African Symposium on Statistical Development (ASSD) is held in Kigali, Rwanda, 15-18 January (United Nations Economic and Social Council. Economic Commission for Africa. African Centre for Statistics [ACS], 2009b).

**2007** – At the IASSIST 2007 Conference In Montreal, Canada, 15-18<sup>th</sup> May, delegates from Africa form a nascent African branch of the Association to advocate for the preservation and sharing of social survey data in African countries (Woolfrey, 2007).

**2007** – Africa's NSOs recommend a Draft Charter on Statistics for the Continent, in collaboration with UNECA and the African Development Bank (Business Report, 21 June 2007)

**2007** – A workshop on National Access to Research Data is organised by the South African Department of Science and Technology (DST), from 27-28 September (Gray, 2007)

**2007** – The Third African Symposium on Statistical Development (ASSD) is held in Accra, Ghana from 3 to 7 December (3<sup>rd</sup> Africa Symposium on Statistical Development [ASSD], 2007).

**2008** – The First African Digital Curation Conference is held in Pretoria, on February 12<sup>th</sup>-13<sup>th</sup> (National Research Foundation [NRF], 2008).

**2008** – The African Association of Survey Data Archivists (AASDA) is launched in Cape Town, South Africa 10<sup>th</sup>-12<sup>th</sup> May (African Association of Statistical Data Archivists, 2008)

## Appendix D: African Microdata Access Survey 2009<sup>16</sup>

### 1. Purpose of the survey

Data sharing between government statisticians and the academic community can be mutually beneficial. With regard to official statistics, the public needs to see academic research based on national survey data as an extension and enhancement of official statistical information. This will increase their trust in official data and in the national agencies responsible for its compilation. Research based on reliable national statistics can also feed into government policies. In the process official data collection, academic research and policy formation will benefit. However, this type of research to support sound policy decisions and economic innovation is increasingly seen to depend on access to original microdata files for researchers. It is important, therefore to test ease of access to national survey microdata for repurposing by researchers. An investigation was therefore undertaken to measure the availability of microdata from surveys conducted by African National Statistics Offices (NSOs) for reuse by researchers in the region.

The process of applying for data tested the effectiveness of data user-producer communication channels at these NSOs for the purpose of data discovery. Historically, little interaction has taken place between official data producers in Africa and those interested in using their data for research and policy analysis. This has impacted on the *accessibility* of African data, which is an important dimension of quality in data management in National Statistical Systems (NSS). Regular communication between data producers and user groups aids data discovery by ensuring awareness among data users of what data is available and its appropriate usage. The nature of data user-producer communications also has implications for the *relevance* of African data, which is another core component of quality management examined in this study. Without feedback from the user community, it is difficult for data producers to provide statistics that will fulfil existing data needs.

An obvious indicator of national data producers' commitment to providing accessible, user-friendly and relevant statistics is the extent and effectiveness of their public communication channels related to data dissemination. This microdata access study was therefore undertaken

---

<sup>16</sup> The author presented the results of this survey as a paper entitled "African microdata access survey 2009" at the 57<sup>th</sup> Session of the International Statistical Institute, held in Durban, 16 – 22 August, 2009.

## AFRICAN MICRODATA ACCESS SURVEY

using contact points available on the websites of the NSOs. Websites are the main public channel of communication of national data producers worldwide.

### **2. Survey method**

A request for survey data was sent to email addresses obtained from the websites of each NSO, in the official language of the country to avoid a language bias in the responses. A specific household survey was identified where this detail was available from NSO websites. Otherwise a query regarding the availability of microdata from their household surveys was submitted to the NSO. Where no response was received to email queries, the data request was re-submitted by fax. Where there were problems with listed fax numbers, telephone contact details were used to confirm the fax numbers were current, in order to limit the possibility of outdated contact numbers being the cause of non-response.

NSOs in ten African countries have NADA web-based catalogues installed by the IHSN/ADP project, to promote metadata standards and facilitate online data dissemination at African NSOs. These include Ethiopia, Gambia, Ghana, Lesotho, Liberia, Niger, Nigeria, Senegal, Sierra Leone and Uganda. The microdata request forms provided with the NADA software were completed to request data from these NSOs.

### **3. Survey limitations**

Internet access is a vital component of data sharing, and the availability of websites and functioning email contact points indicates a certain minimum technological level maintained by these institutions. Website and email information and email responses were essential components of the survey in order to assess the accessibility and data discovery/user communication dimensions of quality management at African NSOs with regard to data sharing. For this reason, the original plan was to use only email contact points or web-based forms for this survey. However, problems with outdated website and email details became an obstacle to communication. This resulted in a decision to re-submit the data requests by fax if no response was received to the original data request submitted by email or via online data request forms. This was because, although an assessment of web-access and email communication is relevant in judging data accessibility, the survey was also an attempt to measure the *willingness* of African data producers to liaise with potential data users and to share their data.

Bureaucratic resistance to placing data in the public domain can be an obstacle to effective data exchange on the continent, and this element needed to be explored by circumventing

## AFRICAN MICRODATA ACCESS SURVEY

email and internet problems wherever possible. This approach was hampered to some extent by the fact that many of the fax numbers provided were out of date or did not function. In the end contact details from the United Nations Economic Commission for Africa (UNECA) website were used in a further attempt to contact the NSOs (United Nations Economic Commission for Africa, 2009). The strategy did not assist the survey because the email and fax details from this source did not elicit any responses and it would seem that the information on this website is outdated.

A definite limitation of the study was the decision not to use telephone contact points. A small number of African NSOs do not provide email or fax contact details on their websites but only provide telephone numbers. These however, were not used as primary contact points in the study. It is probable that a better response rate would have resulted from the inclusion of telephone inquiries to NSOs, not necessarily in terms of data acquisition, but with regard to obtaining information on data access policies. However, it was decided only to use telephone contacts for the purposes of confirming that fax numbers were current, and not to request data via this contact point. This decision was partly based on time constraints, but also on the belief that email and fax communication should suffice as contact points for data user-producer communication. Given that many researchers interested in African survey data would be from foreign institutions, expecting the potential data user to include costly and time-consuming international telephone contact with NSOs would make this method of data acquisition unviable. However, telephone points were used to verify fax details, in an attempt to limit technological constraints as a cause of non-response. This was not entirely successful as telephone contact points sometimes did not function, which meant the currency of all email and fax details could not be confirmed.

Although the author works for a survey Data Archive in Africa, and has met several key people in African NSOs, all attempts were made to avoid using personal contacts, so as not to bias the study. The aim of the research was to ascertain the extent of availability of the official African microdata for researchers in Africa generally, and using personal contacts may have led to the circumvention of restrictions put in place for the anonymous researcher.<sup>17</sup>

#### **4. Survey results**

The results of the survey are summarized below and included in figure 1 of this study

---

<sup>17</sup> As it turned out, the researcher inadvertently contacted a former colleague when emailing the Directorate of ICT at the Kenya National Bureau of Statistics (KNBS). Although the request was subsequently dealt with by other staff at the KNBS, it is difficult to ascertain if this played a role in the eventual acquisition of microdata from this NSO.

## AFRICAN MICRODATA ACCESS SURVEY

### **Category 1: No website access:**

NSOs in eight African countries did not have websites or these could not be accessed (Angola, Comoros, Democratic Republic of Congo, Eritrea, Libya, Mali, Somalia and Zimbabwe). Contact fax and email details for all NSOs except the Ministry of National Planning in Somalia were found on the UN Economic Commission for Africa's website (<http://www.uneca.org>) but these yielded no response, possibly because they were out of date (the details for heads of NSOs was recorded as from 2003).

### **Category 2: No contact**

NSOs for the remaining forty-four countries had websites. Contact problems with 21 of these fell into the following categories:

- Websites had no email details, and fax numbers used for queries received no response (Seychelles and Swaziland).
- Emails were returned as undeliverable, and no fax numbers were listed (Benin).
- Emails were returned as undeliverable, and fax numbers provided could not be contacted (Central African Republic, Congo, Gabon, Guinea, Guinea-Bissau, Rwanda and Sudan)
- Email and fax contact details were provided, but emails were returned, and no replies were received to faxed queries (Burundi, Burkina Faso and Djibouti).
- Email and fax contact details were provided and requests emailed but no reply was received to these, and fax numbers were not contactable (Algeria, Chad, Ethiopia, Gambia, Liberia, Sierra Leone, Togo and Zambia).

### **Category 3: No replies**

In twelve cases NSOs were contacted by email and fax but these requests elicited no response (Cameroon, Cape Verde, Côte d'Ivoire, Egypt, Equatorial Guinea, Lesotho, Madagascar, Mauritania, Morocco, Niger, Tanzania and Tunisia). A meeting with the Director of the Statistics Department of the Moroccan High Commissioners Office for Planning (HCP), Mohamed Taamouti, at a conference in August 2009 elicited further information on data sharing at the Moroccan NSO. Mr Taamouti explained (and later confirmed by email) that the NSO does not have a microdata dissemination policy and data requests are treated on a case-by-case basis (Taamouti, 2009a; Taamouti, 2009b). However, this information would not have been available to a researcher attempting to contact the NSO from the general email contact details on the HCP website.

**Category 4: No data sharing policies**

In two cases email and fax contacts were successful but staff of these NSOs provided the information that they did not share their data for research. Staff of the Mozambican Instituto Nacional de Estatística (INE) provided the information that the institution currently had no microdata access policy that enabled them to supply microdata files, although they had the technology available to place the data in the public domain (the NADA software) (Mozambique). The INE in São Tomé e Príncipe provided the information that the NSO does not supply microdata for research purposes (São Tomé). Discussions with the President of the Mozambican INE at a conference in August 2009 revealed that the INE did, in fact have a policy of providing their datasets to the Mozambican Department of Science and Technology to make available to researchers (Loureiro, 2009). However, this information was not readily available from the INE's website or mentioned to the researcher by the staff at the INE when inquiries were made.

**Category 5: No follow-up**

Staff of three African NSOs responded to data requests but these were not followed up thereafter (Malawi, Namibia and Uganda). Staff at the National Statistical Office of Malawi agreed in principle to supply the data, but then ran into complications regarding how to send the data to the researcher, after which the trail ran cold, despite further attempts to contact them by email. Staff at the Namibia Central Bureau of Statistics made initial inquiries regarding the nature of the data required, but failed to respond further. The Uganda Bureau of Statistics provided the information that the requested dataset would need to be obtained from the Office of the Presidency. Another dataset was requested but there has to date been no response to this.

**Category 6: Onerous access requirements**

The Mauritian Central Statistics Office required payment of 4000 Rupees (\$130/R986) and a visit to the country to swear an oath of secrecy before the Director of Statistics, or the appointment of an in-country representative to undertake this on the researcher's behalf. These onerous conditions hamper access to their data for researchers (Mauritius).

**Category 7: Data provision on payment**

NSOs of a further two countries charged for data access (Ghana and Senegal). The Ghana Statistical Service required payment to obtain microdata files. The amount was \$320 (R2,430) for the microdata from a panel study in three waves, from 1997 to 2006). This is not costly for

the amount of data provided, but is an obstacle to access for researchers from smaller, less well-funded research institutions on the continent. During July the NADA web-based data management software was installed at the NSO in Senegal, and a NADA data request form was emailed and faxed to the NASD (29 July 2009) who responded on 31 July 2009 with an invoice for 411 331 FCFA (\$902/R6,842) for the data from their 2005 poverty survey. This amount would be a definite obstacle to access for most researchers in the region.

**Category 8: Good data provision**

Finally, four NSOs provided relatively obstacle free access to the microdata from their national surveys (Botswana, Kenya, Nigeria and South Africa). The Botswana Central Statistics Office (CSO) and the Kenya National Bureau of Statistics (KNBS) emailed the data files and supporting documentation in response to requests via email (CSO) and fax (KNBS). The Nigerian Bureau of Statistics (NBS) and Statistics South Africa (StatsSA) allow microdata access via their websites, the NBS through their NADA web-based catalogue and StatsSA via their online NESSTAR server. Microdata files and metadata from national surveys were downloaded via these.

**5. Comment on findings**

**5.1 Technological and human resource limitations**

The manner in which the survey was conducted – through email and fax contact details provided on NSO websites – makes it difficult to confirm what proportion of the poor response rate was due to technological constraints rather than the lack of data sharing policies and procedures at these NSOs. However in cases where NSO websites were not available, or email addresses did not function, outdated technology and skills shortages would seem the primary cause of poor response.

Problems with technological infrastructures and lack of technical support is further evidenced by the fact that fourteen of the fifty-three countries targeted by the study used free public email accounts as contact points (Gmail, Hotmail and Yahoo accounts) rather than organizational email accounts. This would seem to indicate problems with the reliability of mail-servers at these NSOs. With regard to international data advancement initiatives aimed at NSOs, it would seem that IT staff in some cases are barely maintaining the NSO websites established or upgraded by international projects. While outdated technology plays a role here, irregular website maintenance appears to be due to a lack of staff trained in ICTs, particularly those with web maintenance skills.

## AFRICAN MICRODATA ACCESS SURVEY

NSOs in Africa have a chronic shortage of skilled personnel who are lured away from the government sector to better paid private sector employment. This is particularly the case with IT professionals, and consequently these institutions are unable to maintain and update their websites. Websites are the principle conduits via which data producers are able to interact with the public. Ideally these should be utilised as efficient data discovery and data dissemination channels. This study provides evidence to support the view that technological and skills limitations prevent effective communication between NSOs and data users in the region.

This explanation was supported by information provided by Dr Helder Salvaterra, of the Directorate of Demographic and Social Statistics of the National Institute of Statistics in São Tomé and Príncipe. Dr Salvaterra informed the researcher that he was aware that the information on the INE's website was outdated and explained that the staff member responsible for maintaining the institute's website had left and had not been replaced. Constraints related to a shortage of IT staff meant that even where NSOs were willing to provide their data for specific research projects, the means eluded them. Some NSOs supplied data as compressed files via email but in some cases files were not in a format suitable for sharing. For example, the Malawi National Statistical Office were willing to send the requested data to the researcher but indicated that the files from the requested survey dataset were too large and numerous to be sent via email and could not suggest an alternative means of transmitting the data files.

Further evidence in this regard comes from an examination of the websites of NSOs participating in the IHSN/ADP's project to install data management tools with web-based NADA catalogues at NSOs. Ideally, the NADA software should be used by NSOs as a data dissemination tool as it provides a facility for requesting and accessing survey datasets online. This could streamline and standardize microdata supply in African countries. However, the study showed that, despite NSO staff being provided with software tools, and training in their implementation, problems with data dissemination still persist. For example some NSOs do not use the NADA online data request form (Niger, Uganda) or the links to data request forms are not active (Liberia) Some participating NSOs are still using the generic forms included with the software, rather than adapting them with their own details (Ghana, Lesotho, Mozambique). Often potential data users are required to print out and fax or mail data request forms to the relevant NSO, which limits the efficiency of the NADA product. These examples once again indicate possible problems with technology at the NSOs, and confirm the shortage of ICT skills in these institutions. DataFirst is working with the IHSN to address this with the new version of the NADA software being introduced out this year.

Funding shortages are possibly at the root of technological and human resource obstacles to effective data sharing in these institutions. The information collected cannot definitely attest to this but this explanation was confirmed to some extent by a discussion with staff from the Zimbabwe Central Statistical Office (SCSO) which elicited the explanation that their website was not available because they had not paid their ISP bill. The chronic state of under-funding evident in many African NSOs reflects the low priority many African governments have allocated national statistical production because they do not see the value of empirical data for government planning, and do not routinely make use of this resource for policymaking.

### **5.2 Public disclosure issues and the bureaucratic culture in NSOs**

The poor response rate in this study can be explained to some extent by the lack of technological resources and shortage of appropriate skills in African NSOs. A substantial number of these institutions lack the human and technological resources for the long-term preservation and sharing of data. However, government attitudes have an influence on the poor situation with regard to data sharing in African NSOs. The Mauritian example in this study is an indication that, despite technological infrastructure and know-how, access to data for legitimate research can be hampered by overly-bureaucratic approaches to data sharing on the part of government functionaries.<sup>18</sup>

The bureaucratic culture prevailing in government structures worldwide can be seen to hamper data sharing by African NSOs. Government functionaries working with national statistics are accustomed to providing their data to a select group of policy-makers, donor bodies and researchers and are unfamiliar with a wider service orientation that includes the public in their client base. This leads to a dearth of institutional resources at African NSOs for data sharing. Staff at these institutions can do little without support for data sharing from their Directors, who themselves will need authority invested in them by government ministers to share the data. Unless policy decisions at ministerial level support the placing of national data in the public domain, NSO functionaries will continue to resist making decisions in this regard, for fear of overstepping their responsibilities.

NSOs in Nigeria and South Africa provide anonymised data via their websites. These two countries have more financial resources available to them than most countries in the region.

---

<sup>18</sup> Mauritius is ranked 51 on the Global Economic Forum's *Networked Readiness Index 2008-9*, ahead of both South Africa (52) and Egypt (76). Only Tunisia has a higher ranking in Africa (38).

However, it is contended that support for this activity from incumbent governments is the main driving factor here. At least in the case of South Africa support for knowledge utilisation at ministerial level plays an important part in StatsSA's more open approach to data sharing. Government support for the sharing of research results is evidenced by South Africa being the only African signatory to the 2004 OECD Declaration on Access to Research Data from Public Funding. Government commitment is vital for data usage and data sharing. Tools provided by organisations such as the IHSN mean data sharing is technically feasible, but this needs the necessary government support to become a reality.

## 6. Conclusion

Where contact was made with NSOs, responses indicate that, with few exceptions, currently data provision is either not part of NSO agendas, or is handled on a case-by-case basis. The former situation wastes national resources by preventing reuse of data by researchers for policy analysis to aid better governance, or to provide innovative input for economic growth. The latter can lead to onerous requirements for data access, again restricting usage and limiting the benefits of data repurposing for African countries. Regional and international support and local enthusiasm for data sharing in Africa is hampered by the paucity of technological and human resources in African NSOs. This is partly the result of limited understanding African policymakers have of the value of data usage for planning and economic development. An appropriately trained workforce supported by enabling technologies are necessary conditions for the effective utilisation of national data resources, but these need to be underpinned by sound data usage policies driven by government decision-makers who appreciate the role of information in the development of modern societies, and are committed to using these resources for national growth.

AFRICAN MICRODATA CURATION INDEX

Appendix E: African Microdata Curation Index

COUNTRY	REGION	DATA CURATION POLICIES	DATA CURATION LEGISLATION	DATA CURATION INSTITUTION	DATA QUALITY FRAMEWORK	MICRODATA ACCESS TOOLS	AASDA MEMBER	DATA CURATION TRAINING	MICRODATA ACCESS SURVEY CATEGORY
Algeria	NORTH AFRICA	NSDS not yet approved	the NSS is legislated by Legislative Decree No. 94-03, 1994 (official data)	Office of National Statistics (ONS) <a href="http://www.ons.dz">http://www.ons.dz</a>	MEDSTAT Programme	N/A	No	In-service training at the ONS.	2
Angola	SOUTHERN AFRICA	NSDS (expired)		Institute of National Statistics	GDDS	N/A			1
Benin	WEST AFRICA	NICI, NSDS	Law No. 99-011 of 1999	National Institute of Statistics and Economic Analysis (INSEA) <a href="http://www.insae-bj.org/">http://www.insae-bj.org/</a>	GDDS	N/A	No		2
Botswana	SOUTHERN AFRICA	No NSDS but has Mission Statement	Statistics Act of 1967	Central Statistics Office (CSO) <a href="http://www.cso.gov.bw">http://www.cso.gov.bw</a>	GDDS; DQAF ROSC completed	N/A	No	In-service training at the CSO	8
Burkina Faso	WEST AFRICA	NICI, NSDS	Law No. 012 of 2007 regulating statistical activities	Institute of National Statistics and Demography (INSD) <a href="http://www.insd.bf">http://www.insd.bf</a>	GDDS; DQAF ROSC completed	N/A			2
Burundi	CENTRAL AFRICA	NICI, NSDS expired - new NSDS being designed) Currently developing a	Decree No. 100/033 1990	Institute of Statistics and Economic Studies (ISTEEBU) <a href="http://burundistats.org/">http://burundistats.org/</a>	N/A	N/A	No		2
Cameroon	WEST AFRICA	NICI, NSDS (expired)	Decree No. 2001/100	National Institute of Statistics <a href="http://www.statistics-cameroon.org/">http://www.statistics-cameroon.org/</a>	GDDS	N/A	Yes		3
Cape Verde	WEST AFRICA	NICI, NSDS	Decree No. 9/2000	Institute of National Statistics (INE) <a href="http://www.ine.cv/">http://www.ine.cv/</a>	GDDS	N/A	No		3
Central African Republic	CENTRAL AFRICA			Directorate of Statistics and Economic and Social Studies <a href="http://www.stat-centrafrique.com">http://www.stat-centrafrique.com</a>	GDDS	N/A	No		2

COUNTRY	REGION	DATA CURATION POLICIES	DATA CURATION LEGISLATION	DATA CURATION INSTITUTION	DATA QUALITY FRAMEWORK	MICRODATA ACCESS TOOLS	AASDA MEMBER	DATA CURATION TRAINING	MICRODATA ACCESS SURVEY CATEGORY
Chad	WEST AFRICA	Currently developing a NICI, NSDS expired	Presidential Decree No. 12 116/P.CSM/MFPT/D G, 1978	Institute of National Statistics and Economic and Demographic Studies (INSEED) <a href="http://www.insead-tchad.org/">http://www.insead-tchad.org/</a>	GDDS; DQAF ROSC completed	N/A	No		2
Comoros	SOUTHERN AFRICA	NSDS		Institute of National Statistics (INS) No website found	N/A	N/A	No		1
Congo	CENTRAL AFRICA	NSDS	Decree No. 2003-133	National Centre for Statistics and Economic Studies <a href="http://www.cisee.org/">http://www.cisee.org/</a>	GDDS	N/A	No		2
Côte d'Ivoire	WEST AFRICA	Statistics Strategy document not yet approved	Decree No. 96.975 of 1996	Institute of National Statistics <a href="http://www.ins.ci/">http://www.ins.ci/</a>	GDDS	N/A	No		3
Democratic Republic of Congo	CENTRAL AFRICA			National Institute of Statistics (INS) No website found	GDDS	N/A	No		1
Djibouti	EAST AFRICA	NSDS	Not available on their website	Institute of National Statistics (INS) <a href="http://www.ministere-finances.dj/Statist/index.HTM">http://www.ministere-finances.dj/Statist/index.HTM</a>	N/A	N/A	No		2
Egypt	NORTH AFRICA	Designing a NSDS with support from Statistics Denmark	Decree No.2915 of 1964	Central Agency for Public Mobilization and Statistics (CAPMAS) <a href="http://www.msriintranet.capmas.gov.eg">http://www.msriintranet.capmas.gov.eg</a>	SDDS; DQAF ROSC completed; MEDSTAT Programme	N/A	No	The Cairo Demographic Centre offers short courses in data management	3
Equatorial Guinea	WEST AFRICA	Statistics Strategy document not yet officially approved	Statistics Law No. 3 of 2001	Directorate General of Statistics and National Accounts <a href="http://www.dgecnstat-ge.org/">http://www.dgecnstat-ge.org/</a>	N/A	N/A	No		3
Eritrea	EAST AFRICA	NSDS		National Statistics Office (NSO) Ministry of Finance and Development	N/A	N/A	No		1

COUNTRY	REGION	DATA CURATION POLICIES	DATA CURATION LEGISLATION	DATA CURATION INSTITUTION	DATA QUALITY FRAMEWORK	MICRODATA ACCESS TOOLS	AASDA MEMBER	DATA CURATION TRAINING	MICRODATA ACCESS SURVEY CATEGORY
Ethiopia	EAST AFRICA	NICI, NSDS	Proclamation No. 303/1972	Central Statistical Agency of Ethiopia (CSA) <a href="http://www.csa.gov.et/">http://www.csa.gov.et/</a>	GDDS	NADA ( <a href="http://www.csa.gov.et/nada/">http://www.csa.gov.et/nada/</a> );	Yes	In-service training is undertaken by the CSA	2
Gabon	WEST AFRICA	NSDS not yet officially approved	Le décret n° 00718/PR/MPAT du 31 mai 1983	Directorate General of Statistics and Economic Studies (DGSEE) <a href="http://www.stat-gabon.ga/">http://www.stat-gabon.ga/</a>	GDDS	N/A	No		2
Gambia	WEST AFRICA	NSDS	The Statistics Act, 1972	Gambia Bureau of Statistics (GBOS) <a href="http://www.gbos.gm">http://www.gbos.gm</a>	GDDS; DQAF ROSC completed	NADA <a href="http://www.gbos.gm/nada/">http://www.gbos.gm/nada/</a>	Yes		2
Ghana	WEST AFRICA	NICI, NSDS expired - new NSDS is being designed	The Statistical Service Law (PNDC Law 135) 1985	Ghana Statistical Services (GSS); <a href="http://www.statsghana.gov.gh">http://www.statsghana.gov.gh</a>	GDDS	NADA <a href="http://www.statsghana.gov.gh/nada">http://www.statsghana.gov.gh/nada</a>	Yes		7
Guinea	WEST AFRICA	NICI	Not available on their website	Directorate of National Statistics <a href="http://www.stat-guinee.org/">http://www.stat-guinee.org/</a>	GDDS	N/A	No		2
Guinea-Bissau	WEST AFRICA	No NSDS but data Development part of PRSP	Decree-Law No. 2 / 91, 1984	Institute of National Statistics <a href="http://www.stat-guineebissau.com/">http://www.stat-guineebissau.com/</a>	GDDS	N/A	No		2
Kenya	EAST AFRICA	NSDS	The Statistics Act No. 4 of 2006	National Bureau of Statistics <a href="http://www.cbs-ke.org/">http://www.cbs-ke.org/</a>	GDDS; DQAF ROSC completed	N/A	Yes		8
Lesotho	SOUTHERN AFRICA	Developing a NICI, NSDS expired -new NSDS awaiting adoption	Lesotho Bureau of Statistics Act No. 8 of 2001	Lesotho Bureau of Statistics <a href="http://www.bos.gov.ls/">http://www.bos.gov.ls/</a>	GDDS	NADA ( <a href="http://www.bos.gov.ls/nada/">http://www.bos.gov.ls/nada/</a> );	No		3
Liberia	WEST AFRICA	NSDS is being designed	The National Statistics and Geo-Information Act of 2004	The Liberia Institute of Statistic and Geo-Information Services (LISGIS) <a href="http://www.lisgis.org">http://www.lisgis.org</a>	GDDS	NADA ( <a href="http://www.lisgis.org/nada/">http://www.lisgis.org/nada/</a> );	Yes		2
Libya	NORTH AFRICA				N/A	N/A			1

COUNTRY	REGION	DATA CURATION POLICIES	DATA CURATION LEGISLATION	DATA CURATION INSTITUTION	DATA QUALITY FRAMEWORK	MICRODATA ACCESS TOOLS	AASDA MEMBER	DATA CURATION TRAINING	MICRODATA ACCESS SURVEY CATEGORY
Madagascar	SOUTHERN AFRICA	NSDS awaiting adoption	Not available on their website	The Institute of National Statistics (INSTAT); <a href="http://www.instat.mg/">http://www.instat.mg/</a> . The website is still under construction	GDDS	N/A	No		3
Malawi	SOUTHERN AFRICA	NICI, NSDS	The 1967 Statistics Act	The National Statistical Office of Malawi (NSO) <a href="http://www.nso.malawi.net/">http://www.nso.malawi.net/</a>	GDDS; DQAF ROSC completed	N/A	No		5
Mali	WEST AFRICA	NICI, NSDS	Ordinance No. 04-008/P-RM of 2004	National Directorate of Statistics and Informatics (DNSI) <a href="http://www.dnsi.gov.ml/">http://www.dnsi.gov.ml/</a>	GDDS	N/A	Yes		1
Mauritania	WEST AFRICA	NSDS	Decree No. 90.026/P/CMSN of 1990	Office of National Statistics <a href="http://www.ons.mr/">http://www.ons.mr/</a>	GDDS	N/A	No		3
Mauritius	SOUTHERN AFRICA	NSDS	The Statistics Act of 2000	Central Statistics Office <a href="http://www.gov.mu/porta/site/cso/">http://www.gov.mu/porta/site/cso/</a>	GDDS	N/A	No		6
Morocco	NORTH AFRICA	NSDS not yet completed	Royal Decree No. 532-67, 1967	Directorate of Statistics <a href="http://www.hcp.ma/">http://www.hcp.ma/</a>	SDDS; DQAF ROSC; MEDSTAT Programme	N/A	No		3
Mozambique	SOUTHERN AFRICA	NSDS	Presidential Decree nr. 9/96 Of 1996	Institute of National Statistics (INE) <a href="http://www.ine.mz">http://www.ine.mz</a>	GDDS; DQAF ROSC completed	NADA <a href="http://196.22.54.6/anda/">http://196.22.54.6/anda/</a>	Yes		4
Namibia	SOUTHERN AFRICA	NICI; Statistical plans form part of National Development Plan 2007-2012	Statistics Act, No 66 of 1976	Central Bureau of Statistics (National Planning Commission) <a href="http://www.npc.gov.na/cbs">http://www.npc.gov.na/cbs</a>	GDDS	N/A	No		5

COUNTRY	REGION	DATA CURATION POLICIES	DATA CURATION LEGISLATION	DATA CURATION INSTITUTION	DATA QUALITY FRAMEWORK	MICRODATA ACCESS TOOLS	AASDA MEMBER	DATA CURATION TRAINING	MICRODATA ACCESS SURVEY CATEGORY
Niger	WEST AFRICA	NSDS	Law No. 2004-011	Institute of National Statistics (INS) <a href="http://www.ins.ne/">http://www.ins.ne/</a>	GDDS	NADA ( <a href="http://www.ins.ne/ncdc/">http://www.ins.ne/ncdc/</a> )	Yes		3
Nigeria	WEST AFRICA	NICI, NSDS	The National Statistics Act of 2007 formed the NBS from a merger of the Federal Office of Statistics and the National Data Bank	National Bureau of Statistics <a href="http://www.nigerianstat.gov.ng">http://www.nigerianstat.gov.ng</a>	GDDS	NADA <a href="http://www.nigerianstat.gov.ng/ncdc/">http://www.nigerianstat.gov.ng/ncdc/</a>	Yes	UNICEF workshop on data management for district statisticians 2008	8
Rwanda	CENTRAL AFRICA	NICI, NSDS	Organic Law No. 01/2005	National Institute of Statistics of Rwanda <a href="http://www.statistics.gov.rw/">http://www.statistics.gov.rw/</a>	GDDS	N/A	No		
São Tomé e Príncipe	WEST AFRICA	Statistics Strategy document not yet officially approved	Article 7 no. 1 of Law no. 5/98, 1998	Institute of National Statistics <a href="http://www.ins.st/">http://www.ins.st/</a>	GDDS	N/A	No		4
Senegal	WEST AFRICA	NSDS	Decree No. 2005-435	National Agency of Statistics and Demography <a href="http://www.anstd.sn/">http://www.anstd.sn/</a>	GDDS; DQAF ROISC completed	NADA	No		7
Seychelles	EAST AFRICA	National Statistics Bureau Act, 2005	National Statistics Bureau Act, 2005	National Statistics Bureau (NSB) <a href="http://www.nsb.gov.sc">http://www.nsb.gov.sc</a>	GDDS	N/A	No		2
Sierra Leone	WEST AFRICA	NSDS; Currently developing a NICI	Statistics Act of 2002	Statistics Sierra Leone <a href="http://www.statistics.sl/">http://www.statistics.sl/</a>	GDDS	NADA ( <a href="http://www.survc.a.net/sl/">http://www.survc.a.net/sl/</a> )	No		2
Somalia	EAST AFRICA				N/A		No		1

COUNTRY	REGION	DATA CURATION POLICIES	DATA CURATION LEGISLATION	DATA CURATION INSTITUTION	DATA QUALITY FRAMEWORK	MICRODATA ACCESS TOOLS	AASDA MEMBER	DATA CURATION TRAINING	MICRODATA ACCESS SURVEY CATEGORY
South Africa	SOUTHERN AFRICA	Strategic Plan 2005; Signatory to the 2004 OECD Declaration on Access to Research Data	Statistics Act, 1999	Statistics South Africa <a href="http://www.statssa.gov.za">http://www.statssa.gov.za</a> ; South African Data Archive <a href="http://www.nrf.ac.za/sada">http://www.nrf.ac.za/sada</a> ; DataFirst <a href="http://www.datafirst.uct.ac.za">http://www.datafirst.uct.ac.za</a>	SDDS; DQAF ROSC; DMID	NESSTAR site	Yes	Statistics SA organised a workshop on data sharing in 1998	8
Sudan	EAST AFRICA	NSDS	Statistics Act, 2003 provides legislation for the NSS	Central Bureau of Statistics <a href="http://www.cbs.gov.sd">http://www.cbs.gov.sd</a>	GDDS	N/A	No		2
Swaziland	SOUTHERN AFRICA	NSDS	Statistics Act, 1967	Central Statistical Office <a href="http://www.gov.sz">http://www.gov.sz</a>	GDDS	N/A	No		2
Tanzania	EAST AFRICA	NICI, NSDS	Statistics Act, 2002	National Bureau of Statistics <a href="http://www.nbs.go.tz/">http://www.nbs.go.tz/</a>	GDDS; DQAF ROSC; Socio-economic Database (ISED)	N/A	No	EASTC offers short courses in data management	3
Togo	WEST AFRICA	Developing a NICI		Department of Statistics and National Accounts <a href="http://www.stat-togo.org">http://www.stat-togo.org</a>	GDDS	N/A	No		2
Tunisia	NORTH AFRICA	5 Year Statistics Plan 2002-2006 has expired	INS Statute, 1969	National Statistics Institute (NSI) <a href="http://www.ins.nat.tn">http://www.ins.nat.tn</a>	SDDS; DQAF ROSC; MEDSTAT Programme	N/A		In-service training by the NSI	3
Uganda	EAST AFRICA	NICI, NSDS	Uganda Bureau of Statistics Act No. 2 of 1998.	Uganda Bureau of Statistics <a href="http://www.ubos.org/">www.ubos.org/</a>	GDDS; DQAF ROSC completed	NADA ( <a href="http://www.ubos.org/nada/">http://www.ubos.org/nada/</a> )	Yes	ISAE ran a short course on data management in 2001	5
Zambia	SOUTHERN AFRICA	NSDS expired - new NSDS being designed	Census and Statistics Act of 1964, CAP 127 of the Laws of Zambia	Zambia Central Statistical Office <a href="http://www.zamstats.gov.zm/">http://www.zamstats.gov.zm/</a>	GDDS; DQAF ROSC	N/A	No		2
Zimbabwe	SOUTHERN AFRICA	NSDS	Statistics Act No 6 of 1999	Central Statistical Office <a href="http://www.zimstat.co.zw">http://www.zimstat.co.zw</a>	GDDS	N/A	No		1

University of Cape Town