STRATEGIES FOR PRESERVATION OF WEB-BASED CONTENT: INTELLECTUAL PROPERTY BARRIERS TO BUILDING THE SUDAN WEB ARCHIVE

A thesis submitted to the Graduate College at the University of Khartoum for obtaining the Degree of PhD in Information & Library Science

BY

Gafar Ali Fadol Ibrahim
BA (Hons), (Information & Library Science), (1998)
University of Khartoum

University of Khartoum

Supervisor

Professor Radia Adam Mohamed
Department of Information & Library Science,
Faculty of Arts

April, 2010
Acknowledgements

The supervisor, Professor Radia Adam Mohammed, lecturer at the Department of Information and Library Science at the University of Khartoum, has provided invaluable advice and direction on every aspect of this study, as well as constant moral support. No Ph.D. candidate could have asked for more. She also made many useful suggestions and recommendations. Her keen intuition and innovative instructions add value and privilege to the thesis.

Furthermore, I would like to express my appreciation and gratitude to the internal examiner, Professor Sami Sherif, Dean of Faculty of Engineering and Architecture (formerly Director of Information Technology and Network Administration) at the University of Khartoum for providing precious insights and cheerful welcoming, and to Dr Hamid El-Dood Mahdi Abd El-Rahman, lecturer of English Literature at the University of Al-Neelain for his valuable advices and for having reviewed the language of the text and Dr Ahmed Tawer, who reviewed the grammar of the Arabic abstract. My gratitude also extends to Dr Sabri El-Haj, Head of Information and Library Science Department at the University of Khartoum, Ms Wafaa Mohamed Osman, Director of Library and Documentation Centre at CWTOA, Dr Lennart Bjoneborn, lecturer at the Royal School of Information and Library Science in Denmark, for allowing quote from their dissertation and thesis and the external examiner Professor Ali Mohamed Shummu, Director of the Press and Printed Press Materials National Council, for their kind efforts.

However, my colleagues at GHD Global, especially Debbie Mace (Administrator), Michael Kemp (Manager) and Kate Rabbits (Marketing Manager), who have been endless supportive in arranging for me to have the time to work on this study and providing me with access to office, laboratories, digital library and
database facilities. Thanks to Mr Ibrahim Al Amin, acting-Director of the Sudan National Library, Ms Najuwa Mahmoud, Head of the Digitisation Project at the Sudan National Record Office and to all librarians and archivists who responded to the survey from worldwide online libraries and archives.

Finally, without my family’s assistance and patience, it would not have been possible to complete this document. Their attention and endless endeavour to create calm study environment entitle them to be regarded as full contributors to its production.
STRATEGIES FOR PRESERVATION OF WEB-BASED CONTENT: INTELLECTUAL PROPERTY BARRIERS TO BUILDING THE SUDAN WEB ARCHIVE

A thesis submitted to the Graduate College at the University of Khartoum for obtaining the Degree of PhD in Information & Library Science

By Gafar Ali Fadol Ibrahim

Abstract

Advancement in Web publishing creates new challenges to management of Web-based content as part of the national library holdings. Such content is important but ephemeral and prone to disappearance if no strategies are made to preserve it. Restrictions of intellectual property rights (IPRs) constitute main barriers to its preservation. This study aims to advocate a balance between the information seeking needs and the protection of IPRs for establishment of the Sudan Web Archive. The national Web space at ‘.sd’ domain is chosen as a setting for the empirical investigation.

The survey and action research methods are employed in order to sample, identify and characterise statistical data that is collected, compiled, analysed and collated using SPSS 16.0 software. The survey method is reinforced by two instruments questionnaires and interviews. The action research method is used to find out a solution to the problems of the study, while the Ranganathan’s laws are revisited in the selection and collection development strategies.

This study is placed within the context of Webometrics field as a new core discipline of information and library science. Results indicate that Alpha index measured by Cronbach is at (0.8) for all provided factors. A correlation is found
significant among pair of factors at (0.05). Results of three months of study show that 25% of the Web-based content is ephemeral and findings reached from similar previous researches show that annual rate of content disappearance is at 44% and the average life expectancy of a Website is only 44 days. It is found that the period from the death of the ‘.sd’ domain in mid-2001 and its resurrection in 18th November 2002 a great deal of national Web-based content probably disappeared. These findings indicate that the Web-based content mortality rates are high and the shorter lifecycle of a Website urges the Sudan National Library to adopt establishment of a Web archive, as proposed in this study in order to save the country’s Web space for future access. It is recommended to form a Web legal deposit law to authorise copying for building the collection of the proposed Web Archive.

**Keywords:** National Library - Sudan; the Web - archiving and preservation; the Web - Intellectual property; the Web - Legal deposit - Sudan; Webometrics, science of.
إستراتيجيات حفظ عتوى الويب: حواجز الملكية الفكرية التي تواجه بناء ذاكرة الشوّدان للويب

رسالة مقدمة لانيل درجة دكتوراة الفلسفة في علم المعلومات والكتب

إعداد: جعفر علي فاضل إبراهيم

المستكشف

أدت التطورات الحديثة في مجال النشر على الويب إلى ظهور تغيرات جديدة تواجه إدارة مصادر المكتبة الوطنية كagenta المنشور على الويب بإعتبار قضاء لا يتمزج من مجموعاتها.

فهذا أدى إلى بيع من أكثر المواد هشاشة وعرضة للضياع إذا لم يتم إعداد استراتيجيات لحفظه. وتُشكّل حواجز الملكية الفكرية وحجة رئيسية في هذا النوع من اختوٍї.

هذى البحث إلى إجراء موازنة بين إستراتيجيات استرجاع المعلومات وبين حماية حقوق الملكية الفكرية حتى يمكن بناء ذاكرة الشوّدان للويب. تم اختبار نطاق الشوّدان على الويب كنقطة الارتكاز والتقصي العملي. واستخدام منهجين: هو منهج البحث المسحي ومنهج البحث التطبيقي بغرض جمع، معالجة، جدولة، تحليل العينة وتحديد البيانات الإحصائية ووصفها باستخدام نظام حزم التحليل الإحصائي للدراسات الاجتماعية. وللمنهج المسحي أداتان هما القابلية والإستبان. ووظف المنهج التطبيقي لإنجاز حل لمشكلة الدراسة كما أعيد توظيف قوانين راجمانثان في استراتيجيات اختيار وتحميم المجموعات.

وتُدرج هذه الدراسة في سياق مجال الويبميتريكا كتخصص جديد ضمن تخصصات علم المعلومات والمكتبات. وتُشير النتائج إلى أن مؤشر معامل ألفنـا الاعتمادية بلغ (0.8) لجملة العناصر التي أُختبعت للتحليل الإحصائي. وأن الارتباط بين تلك المتغيرات له دلالة إحصائية بلغت (0.05) وهي توضح تأثر عمليات الحفظ بموجيز الملكية الفكرية. وتُشير النتائج إلى أن نسبة إختفاء عتوى الويب قد بلغت 25% في ثلاثة أشهر وشير نتائج دراسات مماثلة أن نسبة إختفاء ذلك اختوٍї قد بلغت 44% في السنة وأن متوسط دورة حياة موقع الويب هو 44 يوماً فقط. ووجد أن فترة توقف خدمات نطاق الشوّدان على الويب عن العمل في منتصف عام 2001م وجوداته في الخدمة في نهاية عام 2002م رُبما تسبيت في اختفاء نسبة عالية من ذلك التراث القومي الرقمي. إن إرتفاع نسبة إختفاء هذا النوع من اختوٍї
وقصر دورة حياته تدعو متخذي القرار بالكتبة الوطنية إلى بناء ذاكرة السودان للويب، كما هو مقترح في هذه الدراسة. خفظ عتوى النطاق لأغراض الفنادق المستقبلية. وتوصي الدراسة بإعداد قانون حق الإبداع يختص بالويب للسماح بتنمية جماعات الذاكرة.

كلمات مفتاحية: الكتيبة الوطنية - السودان; الويب، أرشيف، حفظ، ذاكرة الويب - السودان; صناعة الختوى; الويب، حقوق الملكية الفكرية; الويبومتر، علم.
# List of Abbreviations and Acronyms

<table>
<thead>
<tr>
<th>Abbr./Acro.</th>
<th>Full form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AARC</td>
<td>Anglo-American Cataloguing Rules</td>
</tr>
<tr>
<td>ADCIA</td>
<td>Australian Digital Content Industry Agenda</td>
</tr>
<tr>
<td>ADI</td>
<td>American Documentation Institute</td>
</tr>
<tr>
<td>AIDS/HIV</td>
<td>Acquired Immune Deficiency Syndrome/Human Immunodeficiency Virus</td>
</tr>
<tr>
<td>ARIPO</td>
<td>African Regional Intellectual Property Organisation</td>
</tr>
<tr>
<td>ARPANET</td>
<td>Advanced Research Projects Agency Network</td>
</tr>
<tr>
<td>ASCII</td>
<td>American Standard Code for Information Interchange</td>
</tr>
<tr>
<td>ASIS</td>
<td>American Society for Information Science</td>
</tr>
<tr>
<td>BIRPI</td>
<td>Bureaux Internationaux reunis pour la protection de la propriete intellectuelle</td>
</tr>
<tr>
<td>BNB</td>
<td>British National Bibliography</td>
</tr>
<tr>
<td>BnF</td>
<td>Bibliotheque nationale de France</td>
</tr>
<tr>
<td>CBD</td>
<td>Convention on Biological Diversity</td>
</tr>
<tr>
<td>ccTLDs</td>
<td>country code Top-Level Domains</td>
</tr>
<tr>
<td>CCELP</td>
<td>Committee on Coordination of the Electronic Library Project</td>
</tr>
<tr>
<td>CENL</td>
<td>Committee of the Conference of European National Libraries</td>
</tr>
<tr>
<td>CERN</td>
<td>European Organisation for Nuclear Research</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>COBRA</td>
<td>Computerised Bibliographic Record Actions</td>
</tr>
<tr>
<td>CRO</td>
<td>Central Records Office</td>
</tr>
<tr>
<td>CSDL</td>
<td>Conseil scientifique du dépôt legal</td>
</tr>
<tr>
<td>CSIC</td>
<td>Centro de Informacion y Documentacion Cientifica</td>
</tr>
<tr>
<td>CWTOA</td>
<td>Commission on the World Trade Affairs</td>
</tr>
<tr>
<td>DAISY</td>
<td>Digital Accessible Information System</td>
</tr>
<tr>
<td>DARPA</td>
<td>Defence Advanced Research Projects Agency</td>
</tr>
<tr>
<td>DCMI</td>
<td>Dublin Core Metadata Initiative</td>
</tr>
<tr>
<td>DCNRC</td>
<td>Documentation Centre of the National Research Centre</td>
</tr>
<tr>
<td>DDA</td>
<td>Doha Development Agenda</td>
</tr>
<tr>
<td>DDC</td>
<td>Dewey Decimal Classification</td>
</tr>
<tr>
<td>DDS</td>
<td>Document delivery services</td>
</tr>
<tr>
<td>DIU</td>
<td>Dams Implementations Unit</td>
</tr>
<tr>
<td>DMCA</td>
<td>Digital Millennium Copyright Act</td>
</tr>
<tr>
<td>DNS</td>
<td>Domain name systems</td>
</tr>
<tr>
<td>DOD</td>
<td>Department of Defence</td>
</tr>
<tr>
<td>DSEP</td>
<td>Deposit systems for electronic publications</td>
</tr>
<tr>
<td>DTB</td>
<td>Digital talking books</td>
</tr>
<tr>
<td>DVD</td>
<td>Digital versatile disc</td>
</tr>
<tr>
<td>EAD</td>
<td>Encoded Archival Description</td>
</tr>
<tr>
<td>eIFL</td>
<td>Electronic Information for Libraries</td>
</tr>
<tr>
<td>EUCEI</td>
<td>European Union Court of First Instance</td>
</tr>
<tr>
<td>FID</td>
<td>Institut International de Documentation</td>
</tr>
<tr>
<td>FIFA</td>
<td>Federation Internationale de Football</td>
</tr>
</tbody>
</table>

x
Association
FNC Federal Networking Council
FOIA Freedom of Information Act
FTP File Transfer Protocol
GATS General Agreement on Trade in Services
GATT General Agreement on Tariffs and Trade
GDP Gross Domestic Product
GIS Geographic information systems
GNP Gross National Product
gTLDs generic Top-Level Domains
HRINFO Human Rights Information
HTML Hypertext Markup Language
HTP Hypertext transfer protocol
IA Internet Archive
IANA Internet Assigned Numbers Authority
ICA International Council on Archives
ICANN Internet Corporation for Assigned Names and Numbers
ICCSDS International Consultative Committee for Space Data Standards
ICTs Information and communication technologies
IIFLA International Federation of Library Associations and Institutions
IGP Internet Governance Project
IIB Institut International de Bibliographie
IICIP Institut international de cooperation intellectuelle de Paris
IIPC International Internet Preservation
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IIS</td>
<td>Internet Information Server</td>
</tr>
<tr>
<td>ILL</td>
<td>Interlibrary loan</td>
</tr>
<tr>
<td>ILS</td>
<td>Integrated library systems</td>
</tr>
<tr>
<td>INA</td>
<td>National Audiovisual Institute</td>
</tr>
<tr>
<td>IPRs</td>
<td>Intellectual Property Rights</td>
</tr>
<tr>
<td>ISBD</td>
<td>International Standard Bibliographic Description</td>
</tr>
<tr>
<td>ISSI</td>
<td>International Society for Scientometrics and Informetrics</td>
</tr>
<tr>
<td>ITU</td>
<td>International Telecommunications Union</td>
</tr>
<tr>
<td>JCLD</td>
<td>Joint Committee on Legal Deposit</td>
</tr>
<tr>
<td>JISC</td>
<td>Joint Information Systems Committee</td>
</tr>
<tr>
<td>LCWA</td>
<td>Library of Congress Web Archive</td>
</tr>
<tr>
<td>LDAP</td>
<td>Legal Deposit Advisory Panel</td>
</tr>
<tr>
<td>LDSRC</td>
<td>Legal Deposit System Research Council</td>
</tr>
<tr>
<td>LIS</td>
<td>Library and information science</td>
</tr>
<tr>
<td>LOCKSS</td>
<td>Lots of Copies Keep Stuff Safe</td>
</tr>
<tr>
<td>LOM</td>
<td>Learning Object Metadata</td>
</tr>
<tr>
<td>METS</td>
<td>Metadata Encoding and Transmission Standard</td>
</tr>
<tr>
<td>MINERVA</td>
<td>Mapping the INternet Electronic Resources Virtual Archive</td>
</tr>
<tr>
<td>MIT</td>
<td>Massachusetts Institute of Technology</td>
</tr>
<tr>
<td>MODS</td>
<td>Metadata Objects Description Schema</td>
</tr>
<tr>
<td>NARA</td>
<td>National Archives and Records Administration</td>
</tr>
<tr>
<td>NASA</td>
<td>National Aeronautics and Space Administration</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>NCRO</td>
<td>National Central Records Office</td>
</tr>
<tr>
<td>NDIIPP</td>
<td>National Digital Information Infrastructure and Preservation Program</td>
</tr>
<tr>
<td>NDL</td>
<td>National Diet Library</td>
</tr>
<tr>
<td>NEDLIB</td>
<td>NEtworked European Deposit LIBrary</td>
</tr>
<tr>
<td>NIC</td>
<td>National Information Centre</td>
</tr>
<tr>
<td>NIN</td>
<td>National Information Network</td>
</tr>
<tr>
<td>NLA</td>
<td>National Library of Agriculture</td>
</tr>
<tr>
<td>NLM</td>
<td>National Library of Medicine</td>
</tr>
<tr>
<td>NLS</td>
<td>National Library Service</td>
</tr>
<tr>
<td>NRO</td>
<td>National Record Office</td>
</tr>
<tr>
<td>NSF</td>
<td>National Science Foundation</td>
</tr>
<tr>
<td>NTC</td>
<td>National Telecommunication Corporation</td>
</tr>
<tr>
<td>NWA</td>
<td>Nordic Web Archive</td>
</tr>
<tr>
<td>OAIS</td>
<td>Open Archive Information System</td>
</tr>
<tr>
<td>OAPI</td>
<td>African Intellectual Property Organisation</td>
</tr>
<tr>
<td>OASIS</td>
<td>Online Archiving and Searching Internet Sources</td>
</tr>
<tr>
<td>OCLC</td>
<td>Online Computer Library Centre</td>
</tr>
<tr>
<td>OECD</td>
<td>Organisation for Economic Co-operation and Development</td>
</tr>
<tr>
<td>OPAC</td>
<td>Online Public Access Catalogue</td>
</tr>
<tr>
<td>OSI</td>
<td>Office for Strategic Initiatives</td>
</tr>
<tr>
<td>PAC</td>
<td>Activity on Preservation and Conservation</td>
</tr>
<tr>
<td>PADI</td>
<td>Preserving Access to Digital Information</td>
</tr>
<tr>
<td>PANDAS</td>
<td>PANDORA Digital Archiving System</td>
</tr>
<tr>
<td>PANDORA</td>
<td>Preserving and Accessing Networked</td>
</tr>
</tbody>
</table>
DOcumentary Resources of Australia

PDF  Portable document format
P2P  Peer-to-peer
R&D  Research and development
RIAA Recording Industry Association of America
RLG  Research Libraries Group
ROSS Reflective Online Searching Skills
RVI  Rift Valley Institute
SCSE Sudan Centre for Census, Statistics and Evaluation
SIS  Sudan Internet Society
SRB  Storage Resource Broker
SSL  Secure Socket Layer
SUDANAR Sudan Open Archive
Sudanet Sudanese Internet Service Company
Sudatel Sudan Telecommunications Company
SWA  Singapore Web Archive
TCP/IP Transmission Control Protocol/Internet Protocol
TEI  Text Encoding Initiative
TIPC Trade and Intellectual Property Court
TK  Traditional knowledge
TPM  Technological protection measures
TRIPS Trade-Related Aspects of Intellectual Property
UBC Universal Bibliographic Control
UCC  User-created content
UCC Universal Copyright Convention
UDC  Universal Decimal Classification
UKWAC  United Kingdom Web Archiving Consortium
UNDP  United Nations Development Programme
UNESCO  United Nations Educational, Scientific and Cultural Organisation
UNICEF  United Nations Children’s Fund
UNIMARC  Universal MAchine Readable Catalogue
UNIVACI  Universal Automatic Computer I
UNUDHR  United Nations Universal Declaration on Human Rights
URIs  Uniform Resource Identifiers
URLs  Uniform Resource Locators
WAIS  Wide Area Information System
WARP  Web ARchiving Project
WCT  WIPO Copyright Treaty
WDL  World Digital Library
WIPO  World Intellectual Property Organisation
WPPT  WIPO Performances and Phonograms Treaty
WSIS  World Summit on Information Society
W3C  World Wide Web Consortium
XML  Extended markup language
## List of Tables

<table>
<thead>
<tr>
<th>Table No.</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>The Sudan Agenda for Enacting Laws for WTO Conformity</td>
<td>205</td>
</tr>
<tr>
<td>4.2</td>
<td>The Categories of the Registered DNs at (.sd)</td>
<td>219</td>
</tr>
<tr>
<td>5.1</td>
<td>The Response Rates of the Total Custodial Institutions in question (A)</td>
<td>356</td>
</tr>
<tr>
<td>5.2</td>
<td>The Response Rates of Question No. (B)</td>
<td>357</td>
</tr>
<tr>
<td>5.3</td>
<td>The Response Rates of Question No. (C)</td>
<td>359</td>
</tr>
<tr>
<td>5.4</td>
<td>The Response Rates of Question No. (D)</td>
<td>360</td>
</tr>
</tbody>
</table>
# List of Figures

<table>
<thead>
<tr>
<th>Figure No.</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Flowchart of the Proposed Sudan Web Archive</td>
<td>232</td>
</tr>
<tr>
<td>4.2</td>
<td>Web Archiving Operations</td>
<td>234</td>
</tr>
<tr>
<td>4.3</td>
<td>The Four Entities of the NLNZ Metadata</td>
<td>253</td>
</tr>
<tr>
<td>5.1</td>
<td>Response Rates by Type of Institutions</td>
<td>357</td>
</tr>
<tr>
<td>5.2</td>
<td>Rates of Reasons for Web Archiving</td>
<td>358</td>
</tr>
<tr>
<td>5.3</td>
<td>Rates of Preservation Laws</td>
<td>359</td>
</tr>
<tr>
<td>5.4</td>
<td>Rates of Web Preservation Issues</td>
<td>360</td>
</tr>
<tr>
<td>5.5</td>
<td>Analysis of Variance (ANOVA)</td>
<td>361</td>
</tr>
<tr>
<td>5.6</td>
<td>Correlations</td>
<td>362</td>
</tr>
<tr>
<td>5.7</td>
<td>Ratio Statistics</td>
<td>363</td>
</tr>
<tr>
<td>5.8</td>
<td>Regression</td>
<td>364</td>
</tr>
<tr>
<td>5.9</td>
<td>Reliability</td>
<td>365</td>
</tr>
<tr>
<td>5.10</td>
<td>T-Test</td>
<td>366</td>
</tr>
</tbody>
</table>
# Table of Contents

Acknowledgements......................................................... ii  
English Abstract......................................................... iii  
Arabic Abstract ......................................................... v  
List of Abbreviations and Acronyms ...... vi  
List of Tables............................................................. xii  
List of Figures............................................................. xiii  

Chapter 1  The Introduction............................................. 1  
1.1 The Problems of the Study ....................... 4  
1.2 Significance of the Study ..................... 5  
1.3 Hypotheses of the Study ....................... 5  
1.4 Objectives of the Study ......................... 6  
1.5 Methods of the Study .......................... 6  
1.6 Scope of the Study ................................. 7  
1.7 Structure of the Study ......................... 8  
1.8 Terminologies of the Study .................. 9  
1.9 Literature Review ............................... 11  
1.9.1 Background................................................. 11  
1.9.1.1 Web Archiving  
Initiatives ..................................................... 16  
1.9.2 The Review Methodology .............. 30  
1.9.3 Researches on the First  
Dependent Variable ......................................... 31  
1.9.4 Researches on the Second  
Dependent Variable ......................................... 39  

Summery .......................................................... 43
Chapter 2 Conceptual Framework

Overview........................................................................................................ 45

2.1 The Concept of Digital Preservation ....................................................... 46

2.2 Background of the Internet................................................................. 52

2.2.2 Background of the Web .............................................................. 57

2.3 Web Opportunities for Libraries ...................................................... 60

2.3.1 Global Access to Information ....................................................... 61

2.3.2 Literacy Development ................................................................. 65

2.3.3 Provision of Web-Braille Services ................................................. 66

2.3.4 DAISY Digital Talking Books ..................................................... 67

2.3.5 Cataloguing Web Resources......................................................... 67

2.3.6 The Digital Publishing ................................................................. 71

2.3.7 Digitisation of Library Materials ................................................... 73

2.3.8 The Rise of Cyberlibraries.......................................................... 77

2.3.9 Physical v. Virtual Library .......................................................... 79

2.3.10 Building the Information Society ................................................ 85

2.3.11 Web-based Content as Cultural Heritage ..................................... 87

2.3.12 Deep Web-based Content........................................................... 91

2.3.13 Content Industry ................................................................. 92

2.4 Barriers of Intellectual Property .................................................... 99

2.4.1 Challenges of Access to Information ............................................ 101

2.4.2 The Industrial Property ............................................................ 107

2.4.2.1 Protection of......................................................................... 108
Chapter 4 Building the Sudan Web Archive

Overview ........................................................................................................ 184

4.1 Internet Demography............................................................... 185

4.2 Preservation Programmes in the Sudan ......................................... 187

4.3 The Sudan National Library................................................. 193

4.3.1 Functions of the National Library ........................................... 195

4.4 The Intellectual Property Administration ....................................... 201

4.4.1 FCLAW................................................................................. 202

4.4.2 RGIP .................................................................................. 203

4.4.3 CWTOA.............................................................................. 203

4.4.4 PPPMNC............................................................................ 204

4.4.5 TIPC.................................................................................... 206

4.5 Proposal for Web Legal Deposit ............................................ 206

4.6 The Preservation Strategies.................................................... 207

4.6.1 Selection ............................................................................. 211

4.6.2 Harvesting ........................................................................... 216

4.6.2.1 Harvesting the Sudan Domain ........................................ 219

4.6.3 Technology Preservation............................................... 223

4.6.4 Emulation ............................................................................ 223

4.6.5 Refreshing ............................................................................ 224

4.6.6 Migration and Conversion .............................................. 224

4.6.7 Long-term Preservation Plan.............................................. 225

4.7 Web Archive Operations.......................................................... 228

4.7.1 Crawlers and Harvesting Tools .............................................. 235
4.7.1.1 Heritrix........................................... 236
4.7.1.2 LOCKSS........................................... 237
4.7.1.3 NEDLIB.......................................... 238
4.7.1.4 HTTrack.......................................... 239
4.7.1.5 DeepArc......................................... 239
4.7.1.6 Wayback ........................................ 240

4.7.2 Applying the Intranet System 241
4.7.3 Applying an Integrated Archiving System 242

4.7.4 Storage Technologies....................... 243
4.7.5 Creating the Metadata ................. 246
4.7.6 Creating the National Bibliography 254

4.7.7 Implementation of the Plan 257
  4.7.7.1 Infrastructure ....................... 258
  4.7.7.2 The Selection Plan.............. 259
  4.7.7.3 Legality of the Preserved Content.... 259
  4.7.7.4 Proprietary Content... 259
  4.7.7.5 The Descriptive Metadata 260
  4.7.7.6 Display and Access...... 260
  4.7.7.7 Maintenance and Weeding 260
  4.7.7.8 User Feedback...................... 261
  4.7.7.9 Preservation...................... 261
  4.7.7.10 Implementation of the Strategies........ 261
  4.7.7.11 Duration.............................. 261
Chapter 5 The Conclusion

Overview .............................................................................. 264

5.1 Interpretation of Findings .............................. 264
5.2 Discussion ................................................................. 267
5.3 Limitations of the Study ............................... 267

Conclusion ................................................................. 268

References

Arabic references ...................................................... 270

English references ................................................ 272

Appendices

Appendix (A) Questionnaire Form and Analysis 351

Appendix (B) Treaties and Conventions 367

Appendix (C) Chronology of Library Technology 371

Appendix (D) Digital Storage Units 373

Appendix (E) Letters of permissions for Quotation 374
CHAPTER 1

THE INTRODUCTION

Web publishing offers great opportunities, but it also creates an unprecedented surfeit of data in an unstable fragile information environment. The formation of national legislation on information in parallel with intellectual property laws has significant impact on the development of content industry. It enables the Sudan National Library to acquire, preserve, and make available the Sudan Web memory. This Chapter aims to provide general orientation for the study.

Digital preservation is increasingly important subject of research, development and discussion. There is a general perception that the preservation of digital material is more problematic than preservation of analogue or other forms. This is more significantly a cause for concern as preservation of content being made available only in digital form increases because there is no hard copy equivalent to be preserved.

The problems surrounding preservation of digital content include technology obsolescence. As formats, software, hardware, and storage media obsolete and degrade, there is a risk that content becomes inaccessible and unusable. This is not only the problem. There are many others, including the ephemeral nature of much digital content. Also Web-based content may change or disappear before it can be captured and preserved.

Increasingly, digital content is not purchased but ‘rented’ through licenced access rather than physical ownership of an information artefact as required by print counterpart. Therefore, the national library loses control over its digital collections. Digital preservation strategies in use or proposed often involve copying and possibly
changing content. That means the copying processes need to be repeated periodically. The rights of content and any associated protocols, software or domain names may belong to a number of different individuals, institutions, companies, and organisations. Such content may be engulfed by technological protection that prevents unauthorised copying and redistribution that may also inhibit or prevent preservation actions.

This study advocates formation of an alternative legal deposit law and proposes establishment of the Sudan Web Archive in order to provide long-term access to Web-based content for the coming national library patrons. However, the former submitted proposal of this thesis entitled: 'The Information Services and Content Management in the Digital Libraries: How to Digitise Library Resources Using the Portable Document Format (PDF) Technology and imagic DB/Text WebPublisher', has been superseded by the current title for several reasons. First, the emergence of new initiatives on digital preservation in general and Web-based content in particular has stimulated the researcher to change the topic for ‘Strategies for Preservation of Web-based Content: Intellectual Property Barriers to Building the Sudan Web Archive’. Second, there is no initiative on preservation of Web-based content has as yet been introduced from any Third World Countries, including the Sudan, in such field. Third, preservation of digital national heritage is a priority and aims to raise awareness and identify challenges to capture the disappearing digital cultural heritage. Fourth, the PDF file format, which the researcher has proposed as file format for digitisation of library and archives items, has been adopted as ISO international standard (19005-1 Document Management-Electronic document file format for long-term preservation-Part 1: Use of PDF 1.4 PDF/A-1) in October 2005 in the same time of the submitted proposal to the University of Khartoum.
The thesis is organised around five conceptual as well as empirical methodological frameworks. The theoretical part begins with Chapter (1) on the Introduction, Chapter (2) on Conceptual Framework, while the empirical methodological part starts by Chapter (3) on Methodology, Chapter (4) on Building the Sudan Web Archive, and Chapter (5) on the Conclusion.

Throughout this study, the preliminary matter pages of the report are numbered using the Roman numerals (i, ii, iii...etc.), while the body, chapters, tables, figures, questionnaire and end matters have been numbered consecutively using the cardinal Arabic numerals (1, 2, 3...etc.). According to the standards of the Faculty of Graduate Studies at the University of Khartoum, the font size used is 14-point courier new for text, 16-point courier new for headings and subsections and 10-point courier new for the table of contents and the footnotes. Quotations have been assigned (0.5) centimetres in indentation. English and Arabic informative abstracts with key descriptors are provided and substantial number of abbreviations and acronyms are listed alphabetically.

Consistently, citation of English references followed the Harvard Referencing System (2007) also known as ‘in-text’ or ‘author-date’ style, while citation of Arabic references followed the footnote numeric style in order to avoid font perplexing. Each author’s last name is alphabetised as an entry but words such as ‘the’, ‘a’, ‘an’ are ignored and languages other than English as well as titles, definitive phrases, Latin words are italicised. A book of ‘How to Write a Thesis’ by Murray (2006) is referred to for the design of this study. At the end of the study there are appendices containing questionnaire form and analysis, laws, treaties, conventions, chronology of library technology, table of units of digital storage and letters of permissions for quotations from unpublished theses.
and dissertations. As argument develop in this thesis, the researcher cannot guarantee that citations referred to herein and quoted from the Web remain accessible due to the nature of rapid change, decay, dynamism of Uniform Resource Locators (URLs) and their associated links. Nonetheless, the research is conducted in British English because the phenomenon investigated is international in its nature and the findings reached may assist in developing worldwide co-operative solution framework.

1.1 Statement of the Problems

The problems of preservation of Web-based content vary in degree and complexity. These problems are categorised as follows:

1.1.1 Technical Problems

1.1.1.1 Fragility and short life of Web-based content;
1.1.1.2 Dynamic nature of Web publishing and obsolescence of software and hardware, i.e. technical dependency;
1.1.1.3 Content overload;
1.1.1.4 Increase of terrorists’ attacks, cybercrimes, viruses, and spread of Spam and rapid growth of born-digital garbage.

1.1.2 Political and Cultural Problems

1.1.2.1 Growing censorship due to deterioration in international foreign relations;
1.1.2.2 Open nature of Web publishing, i.e. lack of governance, administration and quality;
1.1.2.3 Lack of awareness of Web-based content as cultural heritage.

1.1.3 Legal and Economic Problems

1.1.3.1 Shortages of funds and budgets for long-term preservation;
1.1.3.2 Lack or inconsistency of legislation on national information policies;
1.1.3.3 Dominance of proprietary software, hardware and databases, i.e. barriers of intellectual property rights.

1.2 Significance of the study

Preservation of Web-based content is of great cultural and research value. The growing number of Web-based content adds a pressing need for preservation of such content by building a Web Archive within the Sudan National Library and through its techniques and tools in order to cope with this digital flood of content. It is also more important to focus on Web-based content as national legacy and heritage for the future generations.

1.3 Hypotheses of the Study

The study aims to test the following hypotheses:

1.3.1 Web-based content is ephemeral, fragile and prone to loss due to Websites change and disappearance.
1.3.2 The proposed Sudan Web Archive cannot preserve any Web-based content unless both strategies and legislation are prepared.
1.3.3 Copying for preservation of Web-based content is to be restricted because of recent trends of the GATS and TRIPS agreements towards its commercialisation.
1.3.4 Preservation of Web-based content is vulnerable to terrorists’ attacks, Internet backbone destruction, viruses, hacking and piracy.

1.3.5 Preservation of Web-based content including proprietary software, hardware, databases and other surface Web-based content cannot be done due to restrictions of intellectual property rights.

1.3.6 Preservation of Web-based content is affected by the lifecycle of obsolescence and dynamic change of software and hardware.

1.4 Objectives of the Study

The objectives of the study are to:

1.4.1 Introduce new knowledge on preservation of Web-based content.
1.4.2 Answer the gap of knowledge about the phenomenon under study.
1.4.3 Ensure sustainable access to national Web space.
1.4.4 Advocate establishment of a Web legal deposit.
1.4.5 Recommend building the Sudan Web Archive.

1.5 Methods of the Study

The research objectives are investigated employing the survey and action research methods. The survey has two instruments including the questionnaires and interviews.
1.6 Scope of the Study

For the purposes of this study preservation of Web-based content investigated falls within the following scope:

1.6.1 Content created originally in digital form and acquired by a library or repository through purchase, ownership, subscription, transfer or legal deposit.

1.6.2 Content that is created when the information institutions migrate print and manuscript materials into digital form. The term ‘preservation of Web-based content’ refers both to documentation of materials that are created in digital form and never exist in print or analogue form (also called ‘born-digital’ and ‘electronic documents’) and the use of imaging technology to create digital surrogates of analogue materials for preservation and access purposes.
1.7 Structure of the study

CHAPTER 1
THE INTRODUCTION

This Chapter provides an orientation of the importance of the study. It deals with statement of the problems, the hypotheses, scope, significance, structure of the study, terminologies, literature review and precise summery.

CHAPTER 2
CONCEPTUAL FRAMEWORK

The conceptual framework of the study is discussed in this Chapter along with the theoretical framework, concepts of digital preservation, Web opportunities for libraries, content industry, barriers of intellectual property and summery.

CHAPTER 3
METHODOLOGY

The empirical part of the study starts by this Chapter which acts as precursor to the results. It includes rationale for the study, the research methods, procedures and summery.

CHAPTER 4
BUILDING THE SUDAN WEB ARCHIVE

This is the Chapter where the proposed Web legal deposit and the Sudan Web Archive are presented.

CHAPTER 5
THE CONCLUSION
This is the final Chapter of the study, which summarises and interprets the findings and recommends for further works.

1.8 Terminologies of the Study

For lack of definitions of the terms ‘Web archive’ and ‘Web legal deposit’ this study defines them as follows:

1.8.1 Web Archive

A Web-based information storage, retrieval, resource discovery and access system. Its objectives are to provide information services and enable sustainable access to content that is no longer available on the live Web. This system is often built and established by national libraries and archives to preserve the national Web space and other relevant foreign content about the country.

1.8.2 Web Legal Deposit

A law that authorises national libraries and archives to copy, collect, preserve, manage and provide access to Web-based content of cultural, educational, political and economic significance within national and relevant global domains. As opposed to the traditional print legal deposit law, Web legal deposit law is concerned with captures of ephemeral, dynamic, multimedia and rapid change content transmitted over networked environment and encapsulated in proprietary software. It aims to ensure
availability and sovereignty of access to national Web space and associated links to foreign Web-based content.

1.8.3 Webometrics

Thelwall et al. define the term ‘webometrics’ as follows:

‘the study of the quantitative aspects of the construction and use of information resources, structures and technologies on the Web, drawing on bibliometric and informetric approaches’ (Thelwall et al. 2005,p.82).
1.9 LITERATURE REVIEW

Overview

It is common that literature review section comes in Chapter (2) but an exception to this norm is to place it in chapter one (Clare, 2003, p.27) when the problem dealt with is a phenomenon and this paves the way for discussion of the associated theoretical framework. The aim of the literature reviewed herein is to build on relationships of the most significant literature known about the problems of preservation of Web-based content and relevant fields. The objectives of this part of the study are to fill a gap in literature, develop a better model of the dependent variables, establish a baseline for future studies, and demonstrate the utility of a research procedure. This part also aims to make analyses and interpretations in order to identify areas of prior scholarship to prevent duplication of efforts in previous researches conducted on similar or exact problem in order to find out points of strengths and weaknesses. The literature review section is often divided (Garosn, 2002, pp.97-98) into overview, background, review methodology, research on the first and second dependent variables and summary. The coming rigorous review is meant to justify the need for understanding the status of preservation of Web-based content following Garosn’s divisions.

1.9.1 Background

The field of preservation of library materials is as old as the science of library and information science (LIS) itself. The birth of the library science goes back to Ptolemy Philadelphus (ca. 285-247 B.C.) who founded the enumerative bibliography as a discipline in the fourth century BC at the Bibliotheca Alexandria and the first bibliographer
of notes was Callimachus of Cyrene (ca. 310-240 BC) who might have succeeded Zenodotus of Ephesus in the position of chief librarian (Greetham, 1994).

But the actual development of the science of bibliography dates from the ambitious project of Paul Otlet and Henri La Fontaine\(^1\) when established *Institut International de Bibliographie* (IIB), in Brussels in Belgium by close of the nineteenth century. The IIB developed, after the first International Conference on Bibliography in 1895, into the *Institut International de Documentation* (FID). While Europe was much concerned about bibliographic control and problems of classification of knowledge, the United States (US) witnessed different series of developments in information retrieval. Hence the American Documentation Institute (ADI) was established in 1937 and on January 1968 became the American Society for Information Science (ASIS) which led to the shift of the ‘documentation science’ into ‘information science’ (Kent et al. 2003). Bacon’s (1960) model of a ‘science information system’ shows that the connections between science and its documentation have venerable historical origin. His philosophy along with Otlet’s and La Fontaine’s insistence upon cognitive imperatives in regulation of systematic production, collection and conceptual organisation of scientific documents reveal the ancestry and enduring significance of a key element of their analyses, the concept of the epistemic content of scientific documents, and of documents generally.

The important assumption about documents is that their epistemic content (that is their meaning, or the information they either contain or somehow evoke in the mind of a reader) ensures the efficient movement of scientific information to the

\(^1\) They also envisioned the Universal Decimal Classification (UDC) which was first published in 1905 by the IIB and known as ‘Brussels classification.’
relevant sites of knowledge production, in other word, communication of information to users (Frohmann, 2004). But what is LIS? Researchers provided different concepts to the terminology as of Hauke (2005) who argues that LIS is not only confined to the physical aspects of managing print documents on shelves, but it also plays a key role in the distribution, protection, management and generation of theories of information and knowledge. Though it is rare to find a comprehensive LIS definition but Vakkari has developed a definition of the term ‘library and information science’ as follows:

‘(...) a discipline that views information processes from an information seeking perspective. This does not mean that the research exclusively focuses on information seeking, but this perspective essentially structures the discipline. The objective of the investigation is the information seeking of individuals and groups, the factors that generate this activity, as well as various arrangements and conditions that support the information seeking and provide access to information’ (Vakkari, 1996, pp.25-36).

The researcher infers one core philosophy from the above definition. Throughout the course of human history, libraries, as information retrieval systems, met the information seeking needs through preservation and provision of access to print-based content where information seeking behaviour is marked by characteristics of physicality and static geographic location. Lancaster and Warner (1993, p.13), provide that these information systems aim to match information needs against information messages. The researcher observes that the Web has transformed such characteristics by provision of
virtual dynamic information seeking (or Web searching or browsing) features unlimited by physicality, content carriers and political boundaries. But information seekers often recall information precisely and reliably from libraries than find it out on the Web. In this sense again Lancaster and Warner (1993 *op. cit.*, p.43) comment that both recall and precision ensure that most of the relevant information has been retrieved and that most of the information retrieved is relevant. In contrast, the researcher argues, the Web contains information that transcends the library capacities but usually frustrates information retrieval due to several reasons. It provides abundance of information that causes a new phenomena known generally as ‘information overload’. Furthermore, Web search often displays unwanted results or poor quality of information or a message of ‘Error 404, 503 Page or Document Not Found!’ because content no longer available or URL changed or server disconnected or due to linguistic barriers. So, library techniques employed to curb the phenomenon of the ‘explosion of information’ emerged in the 1960s in print form increasingly used to curb rapid disappearance of content on the Web.

LIS studies of mid 1990s have focused extensive efforts to merge approaches of other disciplines such as computer science and Web studies into its curriculum and accordingly hybrid solutions are directed to address the Web-based information seeking problems. As a result, new trend of LIS study have emerged known as Webometrics² (sometimes called Webometry or Cybermetrics).

Two scholars namely Almind and Ingwersen (1997) coined the Webometrics discipline. Based on the definition provided earlier in section (1.8.3), Web

² This science has international society (International Society for Scientometrics and Informetrics, ISSI), journal and holds annual conferences.
archiving is placed in the context of Webometrics. It is worth to mention that the bibliometric and informetric approaches existed before the emergence of the Web. Several information specialists applied these approaches (Bar-Ilan & Peritz, 2002) in order to undergo systematic evaluation and analysis to help them understand how this medium works, grows, and changes, and how it influences their business and research. New approaches in bibliometric and informetric can provide an appropriate means towards achieving the above goals, and towards establishing a sound theory. The function of the proposed Sudan Web Archive (cf. Chapter 4) is to maintain sustainable access to the national Web space through implementing techniques of Webometrics.

1.9.1.1 Web Archiving Initiatives

During the last decade of the twentieth century a number of international efforts have arisen for developing data management and curation systems for reference and resource collections. For instance, the International Consultative Committee for Space Data Standards (ICCSDS) created an archive reference model and service categories for the intermediate and long-term storage of digital data relevant to space missions. This effort produced the Open Archival Information System (OAIS), which is adopted as the ‘de facto’ standard for building digital archives, and provided evidence that a community-focused activity can have much broader impact than originally intended (NSF, 2007).

1.9.1.1.1 The UNESCO Memory of the World

This Memory of the World programme is established according to the UNESCO Charter on the 'Preservation of the Digital Heritage' adopted at the 32nd session of the General Conference in Paris, France, 17 October 2003 (UNESCO, 2003). Considering that the disappearance of heritage in whatever form constitutes an impoverishment of the heritage of all nations. The UNESCO Preamble states that:

'Recalling that the Constitution of UNECSO provides that the Organization will maintain, increase, and diffuse knowledge, by assuring the conservation and protection of the world's inheritance of books, works of art and monuments of history and science, that its 'Information

---

3 Really existing although not legally stated to exist (Longmann Dictionary of Contemporary English, online).
4 United Nations Educational, Scientific and Cultural Organisation.
for All’ Programme provides a platform for discussions and action on information policies and the safeguarding of recorded knowledge, and that its ‘Memory of the World’ Programme aims to ensure the preservation and universal accessibility of the world’s documentary heritage. (…) Understanding that this digital heritage is at risk of being lost and that its preservation for the benefit of present and future generations is an urgent issue of worldwide concern’ (UNESCO, 2003, op. cit., p.5).

As stated in the above Charter, digital heritage, which is disseminated on the Web, is prone to risks of loss and it is clear that there is an urgent need to preserve such heritage for future generations.

1.9.1.1.2 IFLA Initiative

At the end of the twentieth century a number of trends in preservation were noticeable. In 1986, the International Federation of Library Associations and Institutions (IFLA) launched core Activity on Preservation and Conservation (PAC) in Vienna during a conference on Preservation of Library Materials sponsored by the Conference of Directors of National Libraries with IFLA and UNESCO (IFLA, 2006).

It is evident that Raseroka (2003), former IFLA president, refers to IFLA’s position on access to information, that to address the information divide it is necessary to create in libraries and information centres an environment that allows for free and fair access to information as well as freedom of expression and participation in the knowledge society.
The Federation is coordinating worldwide national library efforts in preservation of digital information including Web archiving (IFLA, 2006, op. cit.).

1.9.1.1.3 The Internet Archive

Researches (IIPC Access Working Group, 2006) forecast that archives of materials published on the Web are still in their infancy, the oldest archive, the Internet Archive (IA), is merely ten years old.

A report prepared by the National Science Foundation (NSF) and the Library of Congress (2003), provides that the project is established in 1996, in San Francisco in the US by the digital librarian Brewster Kahle.

The Archive aims to preserve content disseminated on the Web. In seven years it has developed the largest collection of Web pages in the world - about ten billion Web pages, including 200 million pages on 2000 US Election and 500 million pages related to the terrorist attacks of September 11th, 2001. It is reported that it is vulnerable to accidents, degrading storage media, and changing data formats (Michele, 2004). There is only one backup copy of its collections at the Bibliotheca Alexandria (Cerf, 2005, op. cit., p.xvii)

As a public non-profit organisation without a predictable, steady flow of resources, it seeks stable institutional partners to collaborate in its long-term preservation endeavours. It is the largest public Web archive, which comprises of 40 billion pages, 35 million sites from 1996 to 2004. Recently, the Archive added other content such as video, music and text collections all available online for free public access. It operates on the basis of maintaining snapshot of
every domain per two months of the public Web and its content represent 21 languages worldwide.

The Web is a medium of its own and a Web archive has both the character of library and archive (IIPC Access Working Group, 2006, op. cit.).

1.9.1.1.4 IIPC Initiative

In response to the growing concern of the urgent need of international collaboration for preservation of Web-based content for future generations, the International Internet Preservation Consortium (IIPC) has been formed in Paris in July 2003 (Lupovici, 2005).

Led by the Bibliothèque nationale de France (BnF), (National Library of France), the Consortium also comprises National libraries of Australia, Canada, Denmark, Finland, Iceland, Italy, Norway, Sweden, The UK, The Library of Congress and the Internet Archive (IIPC, 2007). It is noteworthy that under developed countries in Latin America, Asia and Africa, including the Sudan, have no representation to the Consortium.

The IIPC has identified a number of key objectives, which guide and shape its mission. These include:

- Collaborative working, within each country’s legislative framework, to identify, develop and facilitate implementation of solutions for selection, collection, preservation and provision of access to Web content;
- Facilitating international coverage of Web content archive collections within national legal frameworks and in accordance with individual national collection development policies;
- International advocacy for initiatives that encourage the collection, preservation and
access to Web resources (The BnF, 2006, op. cit.).

To achieve these objectives, the Consortium provides a forum for sharing knowledge about Web archiving both within the Consortium and beyond, develops and recommend standards, interoperable tools and techniques to acquire, archive and provide access to Web resources, raises awareness of Web preservation issues and initiatives through conferences, workshops, training events, and publications (IIPC, 2007, op. cit.).

The detailed work of the Consortium is carried out through working groups to define policy, requirements, methods, standards and tools for Web archiving. By this means projects will be developed and defined and could successfully lead to the creation and provision of the necessary tools to fulfil the vision of universal coverage of the Web archive (Lupovici, 2005, op. cit.).

To achieve its mission, the Consortium works to accomplish several goals first, enabling the collection of a rich body of Web-based content from around the world to be preserved in a way that it can be archived, secured and accessed over time. Second, to foster the development and use of common tools, techniques and standards that enables the creation of international archives. Third, encourages and supports national libraries everywhere to address problems of Web archiving and preservation (IIPC Website).

In July 2003, the Consortium is formally chartered at the BnF with twelve participating institutions. The members agreed jointly to fund and participate in projects and working groups to accomplish its goals. The initial agreement has been in effect for three years, and membership used to be limited to chartered institutions. Currently, the Consortium is open to libraries,
archives, museums and cultural heritage institutions everywhere (Lupovici, 2007, p.3).

1.9.1.1.5 The MINERVA Web Archive

The acronym ‘MINERVA’ stands for ‘Mapping the INternet Electronic Resources Virtual Archive’. The project is established in 2000 as Web Preservation Project to launch an initiative to collect and preserve primary source materials. A professional team on different disciplines in the Library of Congress representing cataloguing, legal, public services, and technology services studies methods to evaluate, select, collect, catalogue, provide access to, and preserve Web-based content for future researches. Its mission is to make its born digital material available and useful to the Congress and the American people and to sustain and preserve a universal collection of knowledge and creativity for future access (recently, it is renamed into Library of Congress Web Archive (LCWA), (Ammen, 2002, p.100).

1.9.1.1.6 NDIIPP Initiative

In December 2000, the US Congress passed a legislation mandating and authorising the Library of Congress to develop, execute and lead a National Digital Information Infrastructure and Preservation Program (NDIIPP). The legislation offered USD$100 million for the programme and released USD$5 million in initial funding (Campbell, 2006).

The Programme plan: ‘Preserving Our Digital Heritage: Plan for the National Digital Information Infrastructure and Preservation Program’, outlines a method to implement the preservation process. It is purpose is to develop a national strategy to collect, archive, and
preserve the burgeoning amounts of digital content, especially materials that are created only in digital formats, for current and future reference services. The initial emphasis of the Programme focuses on building a national network of partners, developing and testing digital preservation architecture, and supporting advance digital preservation research. Communication with all stakeholders about the Programme activities is also a key emphasis (Smith, 2006).

The Library of Congress through building partnerships, implements the NDIIPP programme with key federal agencies including the National Archives and Records Administration (NARA), the National Library of Medicine (NLM), and the National Library of Agriculture (NLA), in addition to information technology companies, academic institutions, non-profit entities such as RLG5, OCLC6, and CLIR7, and other stakeholders (The Library of Congress, 2002).

In 2004, a formal network of NDIIPP Digital Preservation Partners is established consisting of eight consortia, each of whom have multiple institutional members. Using a combination of the NDIIPP investments and local matching resources, each consortium identifies and preserves at-risk digital content, develops scalable digital collection and preservation strategies, explores protocols and standards to support partnership operations, and supports the development of tools, models and methods for preservation. The Office for Strategic Initiatives (OSI) at the Library of Congress set up a strategic plan for 2008-2013 for increasing the Library’s digital holdings (The Library of Congress-OSI, 2008-2013).

---

5 Research Libraries Group.
6 Online Computerised Library Centre.
7 Council on Library and Information Resources.
The project explores content selection and collection strategies, probes intellectual property issues related to preservation, considers economic sustainability for digital preservation activities, and addresses a broad range of technical and architecture matters. All participants will also engage in collaborative interactions to identify and share best practices and develop effective partnership network policies and procedures (Smith, 2006, op. cit.).

1.9.1.1.7 UK Web Archiving Consortium

A number of major UK institutions have joined together to form the UK Web Archiving Consortium (UKWAC) to share the costs, expertise, and facilities required to archive selected Websites for future access. These institutions, which form the Consortium, believe that there is enough consistency in the business requirements of Web archiving to maintain an operational environment. It is founded in 2003 to address the technical, legal, collection development, operational, management and financial challenges involved (UK Web Archiving Consortium, 2006).

The Consortium consists of the British Library (Leader partner), the National Archives, the National Library of Wales, the National Library of Scotland, the Joint Information Systems Committee (JISC), and the Welcome Library. Each consortium member selects and captures content relevant to its subject and or domain. For example, the British Library preserves content that reflects national culture and events of historical importance. These could include content focusing on key events in national life, e-thesis, selected blogs, and others to support research material. For content harvesting, the Consortium uses the PANDAS crawler (software) that is applied by PANDORA Web Archive. To ensure longevity, the Archive copies files to different
formats so that they remain accessible as hardware and software changes in the future (Beresfrd, 2006).

Its objective is to collect Websites that were no longer available at their original URL or are missing from the ‘live’ Web when checked by UKWAC. It selectively archives Websites from the UK Web space by prioritising material of research value across the spectrum of knowledge; sites which are representative of British cultural heritage in all its diversity and a small number of Websites which demonstrate Web innovation. The Consortium is preparing for forthcoming secondary legislation associated with the Legal Deposit Libraries Act 2003 and the extension of legal deposit to non-print materials including Websites. In September 2005, seven sites were reported as ‘404’ while thirteen Websites were unavailable for other reasons. In September 2006 the UKWAC contained 1,603 titles and 5,344 individual Websites (UK Web Archiving Consortium, 2006, op. cit.).

In terms of coping with intellectual property rights, the Archive holds Web-based content after obtaining permissions from the copyright holders and refers users who seek to copy the archived material to the rightsholders themselves (Tuck, 2006).

1.9.1.1.8 PANDORA\textsuperscript{8} Web Archive

The National Library of Australia has been collecting, managing and providing public access to significant Australian Web publications since 1996 (Gatenby, 2002). The PANDORA Archive is built in the Library in collaboration with nine partners including Australian Institute of Aboriginal and Torres Strait Islander Studies,

\textsuperscript{8} Preserving and Accessing Networked DOccumentary Resources of Australia.

Each partner has published its own selection guidelines, which define which portion of the Australian Web domain the partner will take responsibility for. In October 2004, it contains 7,000 titles, of which approximately 150 titles have access restrictions applied to them. Its collection is heavily used even though most of its titles are still available on the publisher’s Websites. During the 2003-04 financial years, there were almost 4,300,000 page reviews, an average of 358,326 per month. About 52% of the use is from overseas, 27% is from Australia, and the remainder is from unknown sources. Most of the use, about 60%, is referred by search engines and that these users come to PANDORA not knowing what the Archive is or whether they want to be there. About 18% of users are being referred by ‘expert’ services, such as the Library’s own services including its catalogue, indexing and abstracting agencies. As of March 2007, the Archive contains 14,275 titles, 29,463 instances, and 50% of government publications total of 1.39 terabytes. Usage statistics in the period of 2005 to 2006 are 8,794,547 page views. The Archive is recognised as a reliable and well-organised source for high quality Australian content. It was placed on the Memory of the World Australian Register in August 2004. For ensuring quality of the collection management the Archive maintains the ‘look and feel’ of the original content, which remain persistently without change (PANDORA Website).
The Archive also uses a Web-based archiving system known as PANDORA Digital Archiving System (PANDAS), which is developed by the Library and built in-house. The first release of PANDAS took place in June 2001 with the second version being released in August 2002. The library also developed another open source system 'Xing' for harvesting the databases on the deep Web. The public delivery system is also built using Pache/WebOpject/Java and Oracle to provide resource discovery, navigation and access control services. The actual items of the digital content are delivered as static content through Apache. The service is hosted on Sun Solaris server (Koerbin, 2006).

Most of its content is re-gathered on a regular basis, and the Archive contains 14,155 of these re-gatherings, or instances. The Archive estimates that 7.4% of the whole publications and Websites of its collection are no longer available on the live Web. Many titles have been discarded from the live Websites as they grew and change. A monthly checking for broken links to publisher’s Websites is maintained (Phillips, 2004, op. cit.).

1.9.1.1.9 NWA Web Archive

In August 2000 the Nordic National Libraries applied for, and received a grant from the Nordic Council of Ministers and by the Nordic Governments (Nordunet) to develop an Access Module to the Web Archive. The purpose of the Nordic Web Archive (NWA) project is to develop tools for accessing historical archived collections, including interoperability to enable users to search and navigate the archived Web documents of all the Nordic countries. The NWA strategy is to export the contents of the Web to a common format, and index this output using a search engine, enabling fielded free-text search.
The search engine would provide the possibility to develop Web interfaces for search and navigation, including navigation between different versions of URLs. The project has been completed in July 2002 as a cooperative effort of software developers at every Nordic National Library and a Project Managers at the National Library of Norway (Hallgrimsson & Bang (2003)).

1.9.1.1.10 NEDLIB Web Archive

NEDLIB is short for ‘Networked European Deposit Library,’ which is initiated by the COBRA\textsuperscript{9}-group constituted of eight national libraries, one national archive, two ICT\textsuperscript{10} organisations, and three scientific publishers (Steenbakkers, 2000). The project aims to construct the basic infrastructure upon which a networked European deposit library can be built. The objectives of NEDLIB concur with the mission of national deposit libraries to ensure that electronic publications of the present can be used now and in the future. A permanent standing Committee of the Conference of European National Libraries (CENL) launches it on January 1, 1998, with funding from the European Commission’s Telematics Application Programme. The Koninklijke Bibliotheek (National Library of the Netherlands) leads the project (Manson, 2000).

NEDLIB, aims to develop a common architectural framework and basic tools for building deposit systems for electronic publications (DSEP). It addresses major technical issues confronting national deposit libraries that are in the process of extending their deposit, whether by legal or voluntary means, to digital works (Werf-Davelaar, 1999).

\textsuperscript{9} Computerised Bibliographic Record Actions.
\textsuperscript{10} Information and communication technologies.
The collections of the NEDLIB Web Archive contain 97% of formats such as text or html, text or plain, image or jpeg or imagegif and the remaining 3% constitutes a hurdle for the management (Hakala, 2000, op. cit.).

1.9.1.1.11 WARP Web Archive

The National Diet Library (NDL) in Japan has established the ‘National Diet Library Electronic Library Concept’ in May 1998, which confirms that information resources available on the Web should be collected, archived and provided by the library. In 2000 the Library formulated the ‘Basic Implementation Plan for e-Library Services’ and started experimental projects from 2002. In February 2004 it worked out the ‘National Diet Library Digital Library Medium Term Plan for 2004’ and accordingly has since embarked on projects of constructing digital archives. The Web ARchiving Project (WARP) is the project in which the NDL has selectively collected Web resources to archive and make them available to the public provided that the copyright holders give the library permission to do so. The project started in the financial year 2002 on an experimental basis and moved into the operational stage in the financial year 2006 (The NDL, 2007).

The archived materials have been collected from Websites of the national administrative agencies and institutions, prefecture governments, cities, towns, villages and their consolidation councils before and after consolidation, public-interest corporations or organisations and national universities before their changing into independent agencies, and various events such as the FIFA\textsuperscript{11} World Cup 2002. The Archive has also

\textsuperscript{11} Federation Internationale de Football Association.
collected online periodicals published on the Web by the above-mentioned authorities.

WARP assigns metadata to each Website and online periodical as a unit object to be organised (title) with which different versions (items) collected on different dates are linked using ‘Web-crawler’. The total number of Websites in the Archive is 1,905 in addition to 1,493 online periodicals (WARP Web Archiving Project, 2006).

As of January 2007, total holdings of the Archive are 3,421 titles and 14,358 different items, total number of files is 58.94 million; volume of data is 3,963 gigabytes (The NDL, op. cit.).

1.9.1.1.12 SWA and OASIS Web Archives

The First Phase of Singapore Web Archive (SWA) has been launched on October 18, 2006 with the aim to preserve the ‘.sg’ ccTLD\textsuperscript{12} domain of national and historical significance. Later phases are intended to cover the whole domain though not yet backed by legal deposit (San, 2007).

Another Web Archive called Online Archiving and Searching Internet Sources (OASIS) has been established in December 2005 in South Korea to collect and preserve online national digital heritage and to form standard management policies for the digital resources. The project is funded by the Ministry of Culture and Tourism with USD$1 million with a promise to increase such budget. It is planned to collect one million Web resources by the year 2010. Its collection includes a total of 156,798 with a size of 2.4 terabytes (The National Library of Korea, 2007).

\textsuperscript{12} Country code top-level domain.
The following part describes the methods employed in order to retrieve the relevant literature.

1.9.2 The Review Methodology

This part explains how the researcher went about finding the studies included in the review. For fear of ‘publication bias’ both published and unpublished literature have been synthesised. Within his membership of IFLA, the researcher receives discounted IFLA publications and free quarterly journals and newsletters.

Few print and digital library references are drawn from libraries such as the British Council Library in Khartoum and GHD Digital library in Doha, which is administered by the researcher. Many print items located in the British Council Library concerning books and encyclopaedias of LIS were referred to, while more than 15,000 hits were found in the GHD Digital Library regarding papers, codes, standards, and manuals in preservation and archiving of digital materials in general. Since studies on the phenomenon of disappearance of Web-based content are basically located on the Web, the researcher applied certain search strategies to obtain results on specific keywords. Combinations of the Boolean operators are implemented in order to include and or exclude search results. For example, a search for ‘Internet archiving’ on Google search engine provided 650,000 hits in (0:25) seconds. But when the term ‘Internet’ is combined with ‘Internet or Web archiving’ the search results are reduced to 404,000 hits in (0:15). A search for ‘Web archiving conference’ returned 7,590,000 hits in (0:35) seconds of high value scientific papers. The step is further reduced into 540 empirical studies when searched

13 Gutteridge Haskins & Davey, an intercontinental engineering corporation based on Brisbane, Australia.
IFLANET. Reports of the international organisations, consortia, and Web archives have been looked for directly through visiting their Websites. All studies about preservation or archiving the Web are included, while those providing general digital preservation or archiving sources are excluded in order to limit and narrow studies that focus on the phenomenon under investigation and relate to the six hypotheses and variables of this research.

1.9.3 Researches on the First Dependent Variable

The first dependent variable of this study is the ‘disappearance or loss of Web-based content’ as corresponding to hypotheses stated in sections (1.3.1, 1.3.2 and 1.3.4) mentioned previously. Researches which focus on and support this variable include a Library of Congress Report (2002, op. cit.) entitled ‘Preserving Our Digital Heritage: Plan for the National Digital Information and Infrastructure Program a Collaborative Initiative of the Library of Congress’. The Report warns that the rate of disappearance of Web-based content is high and that 44% of the Websites available on the Web in 1998 are no longer in existence in 1999 and the average life of a Website is only 44 days.

Bailey and Thomson (2006), who work as information specialists at the JISC and Wellcome Library, introduce an article in the D-Lib magazine on ‘Building the UK’s First Public Web Archive’, state seriously, though not providing figures, that there is a danger that invaluable scholarly, cultural and scientific resources are being lost to future generations. The article concludes with investigation of solutions to issues such as, selection, rights management and digital preservation.
Layman (2002) who is a lecturer at the School of Information Management and Systems at the University of California Berkeley, this adds provenance to his contribution, presented a paper entitled: ‘Archiving the World Wide Web’. He raises a question of ‘why archive the Web?’ To answer it, the writer points out that the Web is written in 220 languages (although 78% of it is in English) by authors from every nation. About 90% of Web pages are publicly accessible; a collection fifty times larger than the items collected in the Library of Congress and this makes the Web the information source of first resort for millions of readers. Moreover, he realises that the Web is still in its infancy, and the economic, social, and intellectual innovation it causes is just beginning.

He adds that the Web grows rapidly, adding more than seven million pages daily. At the same time, it is continuously disappearing. He agrees with the Library of Congress’ above report saying that the average life span of a Web page is only 44 days, and 44% of the Websites found in 1998 could not be found in 1999. He asserts that Web pages disappear everyday as their authors revise them or servers are taken out of service, but users become aware of this only when they enter a URL and receive a ‘404–Site Not Found’ message. As ubiquitous as the Web seems to be, it is also ephemeral, and much of today’s Web may disappear by tomorrow if not preserved.

The writer concludes that in the past, important parts of human cultural heritage have been lost because they were not preserved (in part because past generations did not or could not, recognise their historical value, he calls this ‘cultural problem’), also they did not address the storage media (he calls this technical problem), and they did not settle the need for rewards or financial benefits (he calls this economic problem) finally they did not establish legal basis for the
preservation process (and he calls this legal problem).

Each of these problems faces preservation in national libraries nowadays in the case of the Web. Without doubts, Layman’s paper is valuable and it contributes to understanding of the theme of this study and supports both first and second variables. Another valuable magazine article (Rosenthal et al. 2005) addressed part of the technical problem (format obsolescence) entitled: ‘Transparent Format Migration of Preserved Web Content’. Authors of the article are professional librarians at Stanford University Libraries. Its theme is about implementation of format migration for Web content that is transparent to readers and based on content negotiation capabilities of HTTP\(^{14}\) format using LOCKSS digital preservation system (see Chapter (4), section 4.7.2.1.2). This is because without implementing the migration strategy, the content will be inaccessible by the time.

Another important document is a report prepared by Day (2003) entitled: ‘Collecting and Preserving the World Wide Web: A feasibility study undertaken for the JISC and Wellcome Trust’, in 25\(^{th}\) February 2003. The report summarised problems of the preservation of Web-based content as decentralisation, dynamic nature, immature and evolving technologies, and legal challenges. The report also discusses the responsibility of preservation of Web-based content that ranges from national libraries to national archives. It concludes that preservation of this material is possible and feasible though surrounded by difficult challenges.


---

\(^{14}\) Hypertext transfer protocol.
harvesting and guidelines to support preservation’. The author is Assistant Director General of Collection Management at the National Library of Australia. He studies plans to develop understanding of issues related to advancement of collaboration in the areas of Web archiving. He also discusses issues related to the deep content. But he does not discuss problems such as software and hardware obsolescence, economic problem, or legal barriers to collecting and preserving such content.

Three information specialists Ortega et al. (2006) who work at the Internet Lab, Centro de Informacion y Documentacion Cientifica\textsuperscript{15} (CSIC), in Madrid studied evolution of Web content in an article entitled: ‘Longitudinal Study of Contents and Elements in the Scientific Web environment’ in the Journal of Information Science. The study aims to investigate the evolution and state of 738 Websites in two different points in time in 1997 and 2004. It tries to establish the rate of growth and decay of the Web and its elements. The work analyses the most valuable Websites for research that had been published from 1997 until 2004. These Websites are characterised as having a great volume of information and act as an information resource to the scientific community. They have been analysed using the WebMapper 2.0 software of the NetCarta. The study concludes that the percentage of added elements in all cases is very high, confirming the overwhelming increase of information within the period of the study. About 17\% of the Websites have changed their URL over the period of the study. The percentage of vanished Web content is 75.22\%, which represents a high rate of loss during the course of their study. Though this kind of study is valuable for measuring the rate of change and disappearance of the Web-based content,

\textsuperscript{15} Centre for Information and Scientific Documentation.
but it does not suggested a framework or strategy to solve the problem.

Web archiving initiatives, as in his study ‘Long-term preservation of Web content’, in Julien Masanès (ed.) ‘Web archiving’, Day (2005) points out that these initiatives exist to collect ephemeral Web-based content for use by current and future generations of users. To date, most such initiatives have concentrated on the development of strategies and software tools for the collection of such content to provide current access to it through interfaces such as the Archive’s Wayback Machine (see Chapter (4), section 4.7.1.6).

It is noticed that there are no current and up-to-date statistics on Web page longevity, but some studies try to approach this goal. According to Day (2005 op. cit.) estimates show that Web pages disappear on average after 75 days. This rate differs from the Library of Congress’s report as discussed above probably due to duration of investigation.

Koehler (2004), in his article ‘Digital Libraries and World Wide Web sites and page persistence’, in the Information Research Journal, made a longitudinal research on Web page persistence. The research concluded that 33.8% of a sample set of pages selected in December 1996 persisted at their original URLs by May 2003. Studies of longevity of Web references in scientific journals show similar trends. For instance, a research conducted in 2003 by Dellavalle et al. (2003) entitled: ‘Information Science: Going, Going, Gone: Lost Internet References’ in Science magazine reveals that 27 months after publication, the proportion of inactive links to the scientific and medical journals rose to 13%. The exact proportions differ, but similar results have been noted for Web-citations in other biomedical journals, in computer science journals and conferences, and in the informetrics sub-discipline of information science.
The article found that Internet references are frequently gone inaccessible.

A paper presented by Philips (2004, op. cit.) at the International Conference on ‘Archiving Web resources: issues for cultural heritage instructions’, had been planned and held at the National Library of Australia in Canberra from 9th – 11th November 2004, entitled ‘What to collect and how to do it: the National Library of Australia’s selective approach’, describes how online publications and Websites should be collected and preserved. One of the advantages of the selective approach is that Websites, which are inaccessible to harvesting crawlers, can be identified and gathered using other methods. This includes commercial titles, which may require authorisation from the publisher. Major disadvantage of this approach is that it takes a resource out of context and often does not include other resources to which it is linked. Contextual meaning often lost and this will be more critical for some resources and research requirements than others. For copyright reasons, the Archive complied with legal challenges when nine copyright owners asked to remove material because it does not have legal deposit provisions for electronic publications.

A paper presented at the above-mentioned National Library of Australia Conference entitled ‘The Future of Web Resources: Who Decides What Gets Saved and Do They Do It?’ presented by Smith (2004), discusses very essential preservation principles. The writer compares the implications and asymmetry between creation and persistence for collecting institutions. He emphasises that the previous strategies librarians relied upon before for selection and preservation are no longer effective on the Web. He points out that neither preserving everything nor selecting set of high-quality information are entirely feasible though insisting in preserving it in essence. This is
because, according to him, the Web is like the universe contains billions and billions of Web pages most of them invisible to the naked eye. Even the visible pages are often unavailable without the right key, password or authorisation due to economic or privacy barriers. The important fact that Smith reaches is that preservation of Web-based content serves a global, or non-local, demand and thus describes this, in classic economic term as ‘public good’ and that cultural institutions should meet the demand for preserving those public goods. The most significant finding of the paper provides that reliance on legal deposit as enforced on print static material is no longer feasible on the Web environment.

In September 21-22 2006, the 6th International Web Archiving Workshop has been held in Alicante, Spain. A paper presented by Ashenfelder (2006) entitled ‘Web Harvesting and Streaming Media’, concluded that there are several challenges to capturing streaming audio and video files during a Web harvest. His paper examines the main obstacles and focusing on the retrieval of streaming media files via non-HTML\footnote{Hypertext markup language.} protocols, playing streaming media files by means of their metafiles, capturing dynamically generated URLs, and get access for downloading and integrating the files, including accessing and playing the files from within the archived, captured Website. This paper is important for it shades light on challenges of preserving audio and video material.

In the 21st International Conference on ‘Very Large Databases’, September 11-14, 2001 in Rome, Italy, a paper presented by Raghavan and Garcia-Molina (2001), who work at the Computer Science Department at University of Stanford, entitled: ‘Crawling the Hidden Web’, the authors describe straightforward methods for addressing the problem of inaccessibility of deep content. They add that
crawlers only retrieve surface content and ignore search forms and pages that require authorisation or prior registration. This ignored content is of great quality, and value available only in restricted databases. They recommend designing a Web crawler that is capable of extracting deep content from the Web.

The problem is clear for Web content is not subject to gradual decay: it either exists, or it does not. If national libraries and archives fail to construct a new mechanism of preservation, human societies will suffer irreversible damage in their collective social memory (Lusenet, 2005). Paynter and Mason (2006), who work as information specialists at the National Library of New Zealand (NLNZ), discuss the problems of both selective and domain harvesting processes of Web resources in the face of rapid and dynamic networked Webscape. Their paper entitled: ‘Building a Web Curator Tool for The National Library of New Zealand’, to find out solution to the problem, the NLNZ and the British Library initiated a project to design and build a Web curator tool to support selection, harvesting and quality assessment of Web-based content. To address challenges of coverage and comprehensiveness of the collections, Masanes provides an article entitled: ‘Web Archiving Methods and Approaches: A Comparative Study’, published in the Library Trends journal. It discusses the difficulty of sampling and collecting Web-based content and warns that exclusion of any part of the content means loss. Also exclusion of any document linked outside of its domain means incomprehensiveness of the collection process. Masanes wonders about the depth that should be traced and captured in the case of the external links. He affirms that there is no general solution to these challenges, only specific ones based on the goal and policy of the preservation process itself. He concludes that archiving the Web is a
task that no single institution can carry out alone. To investigate the correlation between disappearance and loss of Web resources and the legal problems, the next part deals with the second dependent variable.

1.9.4 Researches on the Second Dependent Variable

The second dependent variable of the study is the ‘intellectual property barriers’ as corresponding to hypotheses suggested in sections (1.3.3, 1.3.5 and 1.3.6) and as tested in Chapter (5), section (5.1). Researches which focus on and support this include a report prepared by Charlesworth (2003) entitled: ‘Legal issues relating to the archiving of Internet resources in the U.K., E.U., U.S.A, and Australia: A study undertaken for the JISC and Wellcome Trust’. The Report concentrates on the legal problems of copyright, content liability, defamatory, obscene or indecent materials. It suggests solution to the problem through enactment of legal deposit in order to avoid direct liability. But it realises that approval of national legislations may take long time. Main result of the study emphasises that there is no coherent international approach to harmonisation of legal rules affecting building of Web archives.

Kavcic-Colic (2002), who presented a paper at the 68th IFLA Council and General Conference, August 18-24, 2002 at Glasgow entitled: ‘Archiving the web-some legal aspects’, discusses the copyright as a legal problem for archiving the Internet. He recommends that national libraries and archival institutions should receive greater governmental legislative support in their striving to preserve the whole written and spoken heritage on all carriers. This would, he suggests, enable these important institutions to uninterruptedly perform their duties in preservation of Web resources. An
important result reached is that implementation of the copyright law on the Web affects access to information.


Of particular relevance in this regard, as of Roos, is the development of the DAISY standard (this will be discussed later in Chapter (2), section (2.3.4) for digitally recorded books for people with print disabilities. Especially audio books would, as the result of the international adoption of this standard, eventually be converted from analogue to digital format. In each case, he goes on, the copyright regime applicable to that particular audio book would require careful consideration and, if necessary, it needs to be revisited. He adds that the DAISY system is not a ‘format’, but rather a standard incorporating different commercial or proprietary formats, together with a degree of encryption capability.

Jacobsen (2007), Head of Legal Deposit and Maps at the National Library of Denmark, presented a paper on ‘Webarchiving Internationally: Interoperability in the Future?’ at the World Library and Information Congress: 73rd IFLA General Conference and Council, 19-23 August 2007, Durban in South Africa, provides that the 2004 Danish Legal Deposit Act enables copying for Web archiving purposes. The biggest challenge, according to him, is to make legal deposit, copyright and other legislation adapt to the Web, so national libraries can preserve its content and make it available to present and future generations.

The paper also describes the challenges the Library has faced in developing its Web Archiving Programme, which includes involvement, as leading partner, in the UKWAC, and membership of the IIPC led by the BnF. Two of the most important findings of the paper are that archiving the Web is both a necessity and challenging task that no institution can achieve alone and the ongoing archiving operation is characterised as positive step. In regard to extension of the legislation to non-print materials, the Library adopts a voluntary approach for Web archiving (this approach would not assist in capture and control of the national Web space as intended because law enforcement is essential) until the necessary regulations are passed which may takes several years. It has started Web archiving in 2001 through the ‘.uk’ domain pilot project. The selection policy is based on Websites deemed to be of historical and cultural significance for the UK To ensure broad coverage, a sample from the Dewey Decimal Classification (DDC) is used for organising the collection. Such move towards American-oriented classification system rather than the Universal Decimal Classification (UDC) it often implements is considered a great shift resulted from the American dominance on the Internet and Web technologies. In terms of the
project cost, the Library commenced a multi-million pound investment in the development of a Digital Object Management programme to address in perpetuity the storage, preservation, access and digital rights management aspects of its digital assets. It is intended that this will be a UK national repository with a projected 300-test bed capacity by 2009. Tuck’s paper addresses the essential needs of Web archiving requirements and it touches most of the ideas covered in this study.

There are many challenges in the race towards building the information society. The international community has been worried about safety and security due to lack of cyberspace legislation, the massive proliferation of spam, cybercrimes, viruses, damage and collapse of networks. Safety and security topics are addressed in a paper by Nash and Peltu (2005) entitled: ‘Rethinking Safety & Security in a Networked World: Reducing Harm by Increasing Cooperation’ at the Conference on ‘Safety and Security in a Networked World: Balancing Cyber-Rights and Responsibilities’, for protection of users and infrastructure from cybercrimes, threats as well as ensuring privacy and security of information. The paper concludes that safety and security measures cannot make a difference unless an international cooperation is maintained. Consequently, any plan for building a Web archive should take into consideration the pre-emptive measures to protect its content from these threats (cf. hacking and piracy in Chapter (2), section (2.4.5)).
Summary

The results reached from the reviewed literature are that though seven million Web pages are added on daily basis as Layman (2002, op. cit.) reports but Web-based content disappear at greater rates. This is supported by the fact that annual Web-based content disappearance rate is 44% and average life of a Website is only 44 days as of the Library of Congress Report (2002, op. cit.) or 75 days as of Day (2005, op. cit.) or 75.22% in seven years as of Ortega et al. (2006, op. cit.) as discussed in researches on the first dependent variable. These results correlate with the hypotheses stated in sections (1.3.1, 1.3.4 and 1.3.6). Causes of disappearance are related to absence or tighter legal restrictions as correlated with all researches on the second dependent variable and the hypotheses stated in sections (1.3.2, 1.3.3 and 1.3.5). Other causes of disappearance range from obsolescence and rapid change of software and hardware to disasters, viruses, hacking and piracy as stated in hypotheses in sections (1.3.4 and 1.3.6). Solving the problem of disappearance requires frequent copying for building the Web archive holdings. The processes of such copying have strong relationship with the legal rights.

It is obvious that the components of the two variables of the literature are relevant to the study. The above-mentioned correlation between copying for preservation and challenges of legal rights illustrates the core of the problem and hence the title of this thesis is constructed to contribute to the understanding of the phenomenon under investigation.

The reviewed literature shows that preservation principles of digital content are no longer the same as analogue. They also show that the principles of the traditional legal deposit legislation of print material are inappropriate for
collecting Web resources. Legislators need to establish a unique legal deposit mechanism as this study recommends in Chapter (4), section (4.5). This part of the study constructs its investigations on the basis of the first and second dependent variables.

The reviewed literature discovered that there is no similar study on Web archiving in the Sudan and that there is no suggested solution to the legal deposit problem, as discussed above. It also noticed that the emergence of several international conferences and workshops on Web archiving began since 2000. This denotes the significance and rising demand for preservation of Web-based content. The literature reveals various gaps as all studies on the second dependent variable focused on the problems of the copyright law only and no solution for the legal deposit is suggested. To sum up this part, it is evident that Web resources have a potential importance because the reviewed literature is likely to add to the existing knowledge. Observably, neither one of the above researches nor any research in the Sudan have investigated the topic of ‘Strategies for preservation of Web-based content: Intellectual Property Barriers to Building the Sudan Web Archive’. The coming Chapter introduces the conceptual framework of the study and is linked to the problems identified in the literature review.
CHAPTER 2
CONCEPTUAL FRAMEWORK

Overview

This Chapter aims to focus on the conceptual and theoretical framework upon which problems of preservation of Web-based content need to be understood. These include definition of the fundamental concepts of preservation of Web-based content mentioned in the literature review in Chapter (1), section (1.9). It also discusses the Web opportunities for libraries, content industry and provides review of the problems of intellectual property barriers in order to gain insights and raise awareness of the importance of the study and to acquire similar experience where appropriate.
2.1 The Concept of Digital Preservation

The concept of digital preservation is broadly located in the context of LIS discipline as justified in the literature review (in Chapter (1), section 1.9).

Disasters stroke library materials long ago before history since the destruction of the Bibliotheca Alexandria by fire and in 1966 when the Arno River breached its banks devastating libraries and museums throughout Florence in Italy. In the twenty first-century, most of library items have been freed from their physical container to be transmitted over networks. As a result, concerns about the longevity of digital content has been summering since the 1980s, though brought into public light by an article that appeared in the Scientific American in 1995 by Jeff Rothenberg entitled: Ensuring the Longevity of Digital Documents’. (Rothenberg, 1995).

Digital preservation of digitised or born-digital content is one of the impacts imposed by the Internet revolution. In building information societies, the production and dissemination of educational, scientific and cultural materials, and the preservation of the digital heritage should be regarded as crucial elements (UNESCO, 2003, op. cit., p.37). An example of digital material that should be preserved is a Microsoft Word document, which was running in 1986 on an IBM computer on an MS-DOS environment and stored on a 5.5 floppy disk (Cohen et al. 2006), or the Doomsday Book of 1086 (NCA, 2005,p.4).

The long-term preservation of information in digital form is one of the most important problems faced by the cultural heritage institutions in the early twenty-first century. Hedstrom has defined digital preservation as:
'The planning, resource allocation, and application of preservation methods and technologies necessary to ensure that digital information of continuing value remains accessible and usable' (Hedstrom, 1996, p.37).

Using this definition, it is clear that the digital preservation problem is not just a technical problem, but an organisational and legal hurdle as well. It may, in fact, be easier to solve many of the technical issues relating to the preservation of digital information than to create organisational and managerial structures to support their consistent application. Hedstrom’s definition also stresses that preservation is about maintaining access to information— not just about the various technical options for long-term storage.

It has been clear for sometimes, from the researcher’s point of view, that preservation of information in digital form requires more than just the preservation of the digital bits and bytes themselves. It has been widely assumed that if digital information is to remain understandable over time, there will be a need to preserve information about the technological and other contexts of digital object’s creation and use. In the past, this was sometimes assumed to mean the concurrent preservation of all the relevant documentation that might be associated with a digital object. At the present time, following other trends in digital library technology, a more sophisticated understanding of this concept is known under the name of ‘preservation metadata’, which will be discussed in Chapter (4), section (4.7.7).

Webb (2004, p.28), is not attempting to define information per se, but argues that digital information is information provided in a digital
form and encoded in discrete bits or binary digits so that the computer recognises it. While it is not surprising that numbers can be represented by bits (after all, digital computing originated with the encoding and processing of numbers), nor that someone found a way of representing text by sequences of bits, one of the achievements of the computer revolution has been the successful encoding of all kinds of information in digital form, including still and moving pictures, sound, graphics, coordinates, colour, and even relationships among data. On the one hand, the ability to encode information in digital form, either by creating it that way (often referred to as being born digital) or by capturing it as a copy in digital form (born again digital) has many profound implications for libraries and archives, and for their users.

On the other hand, the nature of digital information also makes it virtually impossible for humans to extract the intended meaning from what is encoded without the mediation of a computer and the various layers of programming that tell the computer how to interpret the code and represent it in a form that can be understood by humans. To achieve this, digital information is usually stored as a sequence of bits, with other bits that tell a reading system what to do with the code, even where the code starts and finishes. If this is not recognised by available computers and layers of software, it cannot be presented as intended, if at all. Digital preservation or long-term preservation is defined as:

‘the general term for all activities concerning the maintenance and care for [or] curation of digital or electronic objects in relation to both storage and access’ (Verheul, 2006, pp.20-21).
The word ‘digital’ means representation of voice or other information using digits (0 and 1); the digits are transmitted as series of pulses (ITU, 2004, op. cit.). Digital networks allow for higher capacity, greater functionality and improved quality of content. Within digital preservation the main activities can be divided into digital archiving and permanent access. Digital archiving on the first hand means the process of backup and ongoing maintenance of digital objects and the associated software and hardware, as opposed to strategies for digital preservation (Beagrie, 2003).

Also the term ‘digital preservation’ is defined as follows:

‘(...) a set of processes, activities and technology utilized to store and access large amounts of heterogeneous digital data for long periods of time- covering tens if not hundred of years. It enables people or systems to use and understand the data in the far future in spite of the unknown changes in users and technologies such as computer hardware, operating systems and applications’ (Cohen et al. 2006, op. cit., p.9).

Permanent access is usually paired with the term digital preservation, indicating that preservation is only half the battle. Within the digital environment, providing permanent access and adequate rendering of the digital object will be one of the greatest challenges, given the technological changes that have and will continue to occur (CENDI, 2007).

Importantly, the principles of preservation in the digital world are not the same as those of the analogue one, and in essence, define the priorities for extending the useful life of information resources. These principles are longevity, choice,
quality, integrity and accessibility (Conway, 2000).

A significant shift in the terminology of archiving has taken place since 1999. The term ‘electronic’ has been replaced with the word ‘digital’, perhaps indicating a shift from concern about e-journals to the full range of material represented in bits and bytes. While major efforts toward digitising paper materials continue, there is a clear emphasis on objects that are ‘born-digital’. The technical issues of long term preservation are similar once the analogue materials have been digitised, but the fact that there is no analogue original to preserve makes the ‘born-digital’ information all the more fragile. Another significant shift in terminology is the move away from the word ‘archiving’. This term has been problematic from the outset. Those involved with digital information were concerned that ‘archiving’ was too closely identified with records management storage. In addition, the term ‘archive’ has taken on new meanings from e-print and preprint archives, which are primarily repositories with no inherent responsibility for or commitment to long-term preservation. The more common term currently is preservation, which links this activity to the long history of preservation in paper (CENDI & National Agricultural Library, 2004).

The term ‘digital archiving’ is used very differently within sectors. The library and archiving communities often use it interchangeably with digital preservation. Computing professionals tend to use digital archiving to mean the process of backup and ongoing maintenance as opposed to strategies for long-term digital preservation (Beagrie, 2003, op. cit.). It is this latter richer definition, as defined under digital preservation, which has used throughout this study. Another definition of ‘digital preservation’ is provided as follows:
'the managed activities necessary: 1) For the long term maintenance of a byte stream (including metadata) sufficient to reproduce a suitable facsimile of the original document and 2) For the continued accessibility of the document contents through time and changing technology’ (Research Libraries Group, 2002).

All the above definitions have been cited for making the main notion of the term ‘digital preservation’ understandable. What most distinguishes the digital preservation context from the analogue one in place for libraries, archives, museums, and other heritage institutions, is the sheer scale of it. It comprehends vastly larger amounts of information, created in a greater variety of formats, and distributed in new venues to a larger and more heterogeneous user base. The digital preservation infrastructure is characterised by a complex network of relationships and dependencies that are simply unknown in the world of print, analogue sound and image resources (NDIIPP, 2002, op. cit., p.17).

Digital preservation is also divided into two types, first, bit preservation in which the processes used to ensure that the bits comprising a preserved object are not lost or become corrupted over time. Such processes include refreshing, backups, and error correcting code modules. Second, logical preservation the processes used to ensure the understandability and usability of the data, in spite of the unknown changes in technologies and users in the future (Cohen et al. Op Cit. 2006).

The next section provides historical background of the Internet and the Web as incubators of digital content.
2.2 Background of the Internet

To understand the approach of preservation of Web-based content, there is a need to draw lines between both the Internet and the Web. Researches (O’Neill, 1998) look at the Internet as the foundation and access on which the Web architecture is based.

Moreover, the International Telecommunications Union (ITU), provides that the Internet is a global network of interconnected networks that applies the Internet protocol (2004, op. cit.). On October 24, 1995, The Federal Networking Council (FNC), in the US unanimously approved a resolution defining the term. This definition is developed in consultation with members of the Internet and intellectual property rights communities. It is agreed that the ‘Internet’ refers to the global information system that:

- 'is logically linked together by a globally unique address space based on the Internet Protocol (IP) or its subsequent extensions or follow-ons;
- is able to support communications using the Transmission Control Protocol/Internet Protocol (TCP/IP) suite or its subsequent extensions or follow-ons, and/or other IP-compatible protocols; and
- provides, uses or makes accessible, either publicly or privately, high level services layered on the communications and related infrastructure’ (FNC, 1995).

The ITU’s definition focuses on the Internet as a global ‘network’ while the FNC’s definition makes it clear that the ‘Internet’ is a ‘global information system’ which operates on communications and related infrastructure for
provision of public or private high level services. This system has greater impacts on library functions and services and such impacts will be discussed later in this Chapter.

The Internet Governance Project (IGP) adds that the Internet is not hardware standard or a physical infrastructure. It is based on a set of software instructions (known as ‘protocols’) for sending data over networks. The IP works by dividing messages up into ‘packets’ and attaching sender and receiver addresses to those packets so that they can be routed to their destination. A closely related protocol, TCP, governs error control and the rate at which packets are sent. The IP can operate on many different physical technologies, and can be used as the underlying communication device for almost any kind of higher-level software application, such as accessing Websites, word processing, streaming video, voice communication or games. The key concept is ‘internetworking’; these protocols are designed to link networks to networks (IGP, 2004).

The Internet is not a specific software application, it is a carrier that allows software applications to communicate and interoperate. Thus, the IGP defines the ‘Internet’ as:

‘the global data communication system formed by the interconnection of public and private telecommunication networks using Internet Protocol (IP), TCP, and the protocols required to implement IP internetworking on a global scale, such as DNS\textsuperscript{17} and packet routing protocols’ (Loc. cit., p.7).

Early in 1951, the first commercial electronic computer known as UNIVACI\textsuperscript{18} was submitted to the US

\textsuperscript{17} Domain Name System or Service.
\textsuperscript{18} Universal Automatic Computer I.
Bureau of Census. Shortly after computers became available at universities, research projects were conducted to link these computers for information sharing (Sadowsky et al., 2003, p.1).

Beginning with interconnection of machines at four universities (including two branches of University of California) in 1969 under the name ARPANET, the Internet has grown by leaps and become a network of networks of phenomenal proportions (Carl-Mitchell, 1994).

But what is known today as the Internet, as Kent et al. (2003) point out, has its origin in the early 1960s as a project of the US Department of Defence (DOD). In fact, the origins of the Internet go back to 1969, but it was in the early 1990s, with the development of the World Wide Web and graphical browsers, that the Internet really took off as a commercial undertaking (The ITU, 2004, op. cit.).

Some studies (Leiner et al. 2000) assume that early-recorded historical background of the social interactions that could be enabled through networking was a series of memos written by J.C.R. Licklider from Massachusetts Institute of Technology (MIT) in August 1962 who proposed the concept of ‘Galactic Network’. Licklider envisioned a globally interconnected set of computers through which everyone could quickly access data and programmes from any site. In spirit, the concept looked like the Internet of today. In October 1962, he became the first head of the computer research programme at the Defence Advanced Research Projects Agency (DARPA).

Leonard Kleinrock at MIT developed the theory of packet switching that later became the basis for Internet connections. Lawrence G. Roberts, a professor at MIT, connected two computers via telephone lines in 1965. Roberts became part of DARPA in 1966 and developed the plan for Advanced Research Projects Agency Network (ARPANET). Its
specific aim was to give American science and technology a boost in response to the former Soviet Union’s launch of Sputnik, the first artificial satellite and a pre-emptive security facility. DARPA wanted to make a network that was smart enough to spontaneously recover from problems such as power failures, interruptions in communication lines in the case of a nuclear attack. It called its network DARPANET (Wiseman, 2000, p.4).

Another hypothesis claims that the actual Internet finds its origin in Robert Kahn’s idea of open-architecture networking, or ‘Internetting’ (Eck, 2002, p.14). His idea was that an open architecture would be able to connect multiple independent networks, each network itself having a different operating system and design. Such as open-architecture network required a new communication protocol, the TCP, which was designed in 1973-74 by Robert Kahn and Vinton G. Cerf\(^{19}\) (The OECD, 2005) and later split into TCP/IP in 1978 (Cerf, 2005).

The National Science Foundation (NSF) extended its network in the 1980s to increase international educational and research opportunities. This network used NSF computers for data storage, file transfer, and E-mail. NSFnet evolved into what is now known as the Internet. It incorporated the TCP/IP format in the UNIX operating system for data packet transmission (Abraham, 1995). This format was used to connect the original ARPANET, with a packet satellite network linked the US and Europe, and with a ground-based packet radio network, into the first Internet (Wiseman, 2000, op. cit., pp.5-6).

As the Internet grew, there became an increasing need for better tools to access and organise information. Using File Transfer Protocol (FTP)

---

\(^{19}\) Known as the father of the Internet, holds a Ph.D. in computer science from University of California Los Anglos and now Vice President & Internet pioneer at Google, Inc.
information could be shared between computers. The protocol enables a user to obtain files of text, executable programmes, graphic films, sound files, and other information, but early restrictions of FTP sites allowed only short names with little description about content. Files were often named README, INDEX, NOTICE, ABOUT, and other variations. These file names were thousands of FTP sites across the world, but each site was a collection of files that were not searchable. A user would have to visit each FTP site and browse through the files to hopefully discover relevant information (Swain, 1994).

Finding information on the Internet seemed somewhat like trying to find a needle in a ‘haystack’. An added dimension to the ‘haystack’ metaphor is that the Internet environment is a dynamic collection of networks. Archie, Veronica, Gopher, Wide Area Information System (WAIS), Mosaic, AltaVisat, HotBot, NorthernLight, Excite, Lycose, Inforseek, and others were the early search tools of the Internet— the first attempts to provide more order and searchability. Although these early tools are still accessible today, they were developed prior to Web browsers and have generally been replaced by more popular Web search tools such as Yahoo and Google (Valauskas, 1994; Bjorneborn & Ingwersen, 2001).

The proliferation of LIS publications addressing and researching aspects of the Internet still a relatively new phenomenon. In fact it is difficult, if not impossible, to find an information science periodical without one article dealing with Internet technology (Greenberg, 2003).

There is growing controversy concerns the question of ownership of the Internet domain name space. There are two competing arguments as to the ownership. The first is that since DARPA, and later the NSF funded and developed the Internet, the US government owns the Internet domain name space.
Under this line of reasoning, IANA’s\(^{20}\) authority to manage the domain name systems (DNs) derives from its position as a government contractor (Albert et al. 1999). According to ICANN\(^{21}\) (2004) there is growing legal tension between the expansion of the DNs or ccTLDs into generic top-level domains (gTLDs) such as (.aero, .biz, .coop, .info, .museum, .name and .pro) and the business identifiers or trademarks as will be discussed in section (2.4.2.1).

### 2.2.2 Background of the Web

The Worldwide Web (WWW or simply the ‘Web’) is viewed as a content transmission facility that is defined as:

> 'An Internet system that distributes graphical, hyperlinked information, based on the hyperertext transfer protocol (HTTP). The web is a global hypertext system providing access to documents written in a script called Hypertext Markup Language (HTML) that allows its contents to be interlinked, locally and remotely. It is designed by Tim Berners-Lee at the European Organization for Nuclear Research (CERN) in Geneva' (Noruzi, 2004).

---

\(^{20}\) Internet Assigned Numbers Authority, a subsidiary of the Information Sciences Institute (ISI) of the University of California founded in 1970s (Albert et al. 1999, op. cit., p.20) currently operated by ICANN.

\(^{21}\) Internet Corporation for Assigned Names and Numbers.
As per Noruzi’s definition, the Web is a hyperlinked content of information that is transferred on global basis. Diffusion of hypertext content adds new materials and responsibilities to library management as will be discussed later in the coming sections.

The Web is a global, virtual publishing environment. The term is often mistakenly used as a synonym for the Internet itself, but the Web is actually something that is available over the Internet, just like the e-mail and many other Internet services. Tim Berners-Lee, who is now head of the World Wide Web Consortium (W3C), created the first server, browser and editor, the HTML code, the URL address and the HTTP transmission protocol at CERN in 1990. CERN released the Web into the public domain in 1993 (CERN, 2004). But Cerf provides that the Internet has entered the public arena in 1994 (Cerf, 2005, op. cit.). Another person not always noticed, Robert Cailliau, has also contributed with Tim-Berners-Lee to development of the Web (Sadowsky et al. 2003, op. cit., p. 14).

The Web is composed of sites on intranets behind firewalls, fee per access sites, and other instances that require prior authorisation. Each site has a unique 32-bit identifier as its IP address. This address, in turn, is divided into 4 octets of 8 bits each often shown separated by dots for instance (132.174.1.5). Since each octet is 8 bits, it can range in value from 8 to 255, creating more than 4 billion potential addresses in the total address space (O’Neil, 1997).

The broadest interpretation of the Web is a collection of HTTP servers operating on TCP/IP interconnected network. Its narrower interpretation, however, consist of all active HTTP servers that receive, understand, and process client requests. Its accessibility can be determined from the response code returned to the
client who attempts to get connected. Response codes in the 200-299 range or 300-399 range (the same as DDC classification schema) indicate that the server received, understood, and processed the client request. Response codes in the 400-499 range indicate that a client-side error prevented completion of the request; these include codes that indicate access to the server is not authorised or forbidden. Response codes in 500-599 range indicate that a server-side error prevented completion of the request (O’Neil et al. 2003).

Text documents, images, multimedia and many other items of information, referred to as resources or content, are identified by short, unique, global identifiers called Uniform Resource Identifiers (URIs) so that each can be found, accessed and cross-referenced in the simplest possible way. This could be done through breaking down messages into packets by the TCP protocol (Wiseman, 2000, op. cit.). The most important feature of the Web is the e-mail service. The first e-mail system had been devised by Ray Tomilson at DARPA, while the first e-mail that was sent in 1971 between adjacent computers at DARPA headquarters (Cerf, 2005, op. cit., p.xvi).

Web documents are known as hypertext with links to further associated documents, on the model of references in a scientific paper or cross-references in indexing. With digital documents, these cross-references or nodes can be followed by a mouse-click. Entry to Website usually starts at the Home page, which is roughly

---

22 In 1945, Vanevar Bush proposed a machine called a ‘Memix’ that could read and retrieve associative links between different aspects of documents using microfilm. This process was called “associative trails.” Ted Nelson (Theodor Nelson) invented the idea of the hypertext in the mid-1960s but the first recorded instance of the word ‘hypertext’ in print only came on February 3, 1965 (Poole, 2005, pp.186-18).
equivalent to the title page in the print environment. The homepage often provides information about the site, and may also function as a table of contents. Following the homepage, the most essential bibliographic unit on the Web is the Web page (static or interactive HTML file). Simply, Web pages are files lying around on the hundreds of thousands of computers connected to the Internet. The Web page is a distinct entity that is identified by a unique URL. Wandering from one document (Webpage) to another is called browsing. Some people do this just for fun, following links just to see what’s there. This is usually called ‘surfing the Web’ (Stout, 1996, p. 7; O’Neil, 1998, op. cit.). The address of the first Website ever created was ‘info.cern.ch’ and the first Web page address was ‘http://info.cern.ch/hypertext/WWW/TheProject.html’ (CERN Website), while the first library Website ever launched at Virginia Tech University, the University of Michigan and the US Naval Research Library (OCLC, 2007) (see also appendix (C) for chronology of library technology). Sometimes, the Web is looked at as ‘digital library’. But finding information on this library is difficult, for the quality of content varies significantly. As a result, the library services have been challenged by the open nature of the Web publishing (Kuny & Cleveland, 1998, op. cit.).

2.3 Web Opportunities for Libraries

The historical patterns of growth of the telecommunications industry have been associated with the development of the alphabets and writing and the inventions of the telegraph in the 1840s, the telephone in the 1870s, radio telegraphy or ‘wireless’ in the 1890s, radio broadcasting in the 1920s, television broadcasting in the 1950s, geostationary satellite communications in the
1960s, computer communications in the 1970s, optical communications in the 1980s, and the Internet and mobile communications in the 1990s (The ITU, 2004).

The convergence of the ICTs generally and the Web in particular are changing the face of the world and creating more opportunities as well as adding immense challenges. The Web has reshaped the nature of the workplace and the way librarians do their work in libraries, basically it is driving them towards a new outstanding era of an information society a concept which will be dealt with later in this section (Calra, 2005).

2.3.1 Global Access to Information

Reference to the British Library’s strategy (2005-2008), libraries have traditionally existed to collect and organise information, make access to knowledge more freely, and preserve the memory of ideas for future generations.

IFLA Internet Manifesto (2002) recognises that libraries and information services provide essential gateways to the Web. For some they offer convenience, guidance, and assistance, while for others they are the only available access points. They provide a mechanism to overcome the divisions created by discrepancies in resources, technology, and training.

Hart (1992), founder of the Project Gutenberg23, asserts that the greatest value created by computers would not be computing, but would be the storage, retrieval, and searching of what has been stored in our libraries.

---

23 Project Gutenberg is the Internet’s oldest producer of public domain electronic books containing over 17,000 (e-Books or e-Texts). The Project has been a part of celebrations of the 100th Anniversary of Public Libraries, starting in 1995. Its Philosophy is to make information, books and other materials available to the general public (Project Gutenberg Website).
The unique function of the library and no other institution is to carry out this long-term systematic work. Such functions still not changed in the digital world but digital technology changes the balance of power among the core functions (Ershova & Hohlov, 2002). Interlibrary loan (ILL) and document delivery services (DDS), for instance, have been transformed using e-mail file attachments. Reference to OCLC24 (2004), the Online Computer Library Centre, in 2004 daily information exchange using e-mail attachments in the US reached 16.5 billion while ILL accounts for 51,000. In 2000, OCLC has introduced the ILLiad software as a leading ILL management tool that automates routine loan functions and provides sophisticated tracking statistics to library staff (OCLC, 2000,p.4). Similarly, compilation and creation of national bibliographic databases have been conducted, distributed and accessed on the Web as an effective and timeless mean of delivery (Knutsen, 2006).

Most cultural-memory institutions dealing with documentary resources find that digital technology affects them in a number of ways that are particularly important to their fundamental functions and responsibilities. Library Web-based databases are connected to high-level networking and format standards such as Z39.50, EDIFACT25, and UNIMARC26 and there are complex relationships emerge as a result of intellectual property rights (Llion, 1994). Web-based content affects the type of format found in collection, whether at a physical level, such as the presence of CD-ROMs, online publications, hard disks, tapes, or

24 Founded by Frederic Kilgour in 1967 in Ohio.
25 EDIFACT is the recognised international standard for electronic data interchange trading in a wide range of commercial and non-commercial sectors.
26 Universal Machine Readable Catalogue format for storing and exchanging bibliographic data.
at a genre level such as datasets, databases, e-journals, geographic information systems (GIS) products, e-manuscripts, multimedia, e-books, e-records, digitised pictures, sound, e-mail collections, Web pages, and many others. Hence coordination within sectoral and cross-national is the top priority in order to save time and resources (Lunghi, 2006).

The Web has been widely used in libraries since it is invention and it has affected all aspects of their services. It is used as a tool for accessing information resources. For instance, academic, special, public and national libraries have set up Websites for their own or within their mother-institutions aiming to reach wider and global users. But one of the most important impacts on the technical operations is how to organise the diffusion of the Web resources for ease of access as often managed by the traditional libraries (Wang, 2000).

The Web enabled libraries to provide their searchable online catalogues accompanied by help instructions for information literacy and user education about how to use and access its content. Library Websites usually contain links to information about the library, photocopying services, DDS or ILL and lending policies. Most sites have their real and virtual address details as well as sitemaps to their physical location. Feedback or ask a librarian are very essential tools for keeping in touch with their patrons. Everybody can send his or her comments and requests through these electronic facilities and expect immediate response and answer for their questions. But these Websites need to ensure that content remains current and accessible (Titoko, 2005).

It could be said that the Web has changed the fundamental roles, paradigms, and organisational cultures of library and also the role of
librarians (Pascoe, 1996). As a result, those information professionals have been affected by the change in their roles, knowledge, skills and competencies (Roland, 1998).

One of the most important impacts created by the Web is what this study calls the ‘social divide’ in libraries. Due to such trend, library users have abandoned the intimacy of the physical library and increasingly request and deal with digital materials available as digital forms only and there are emerging attitudes towards the fate of both the library and the book as container and incubator of information. The fate of the library, as Frischer (2005) argues, is an epiphenomenon of the fate of the book itself. The good news, Frischer goes on, is that the traditional printed book is doing better than ever. The same digital technology that might seem to threaten the book’s existence is also giving us ‘print on demand’, making it easier and cheaper to produce books, reissue them, and publish new editions, all in relatively small print runs.

The researcher comments that the real problem that librarians may soon face is not the death of the print book but the profusion of new titles, reissues of old publications, and new editions of scholarly books by living authors - all made more economical and practical by print on demand. This means that both print books and digital books are going to coexist together as genuine forms of information in the foreseeable future. This is because digital divides within communities and among societies widely grow due to certain social factors such as poverty, power supply, and illiteracy.
2.3.2 Literacy Development

The reading habit is the aim and essence of all textual material. Screen reading has become immensely widespread and is an integral part of interactive communication and reception of all kinds of digital texts. Users often rely on information literate or reference librarian’s skills in order to learn how to locate, assess, and use information effectively to solve a particular problem or make the right decision. Research heavily supports the general belief that most people still prefer to read lengthy texts on paper (Nair, 2006, p. 7).

Reading, or the process of understanding written linguistic messages, is a complicated cognitive task involving important parts of the brain. It takes years of training to become a fluent reader. Even small disturbances in typography, ergonomics, or word semantic can disrupt the reading process and bring the act of reading to an end. Current used digital equipment and software cannot compete with printed-paper as a medium for sustained and uncomplicated reading (Hillesund, 2005).

Web impacts also transcend the traditional information literacy skills developed as part of library services for academic and long learning programmes to incorporate these skills on the cyberspace. As information seeking behaviour, Bruce (1997) notes that information literacy includes aspects of computer literacy, learning to learn, information skills, information technology literacy, and library skills. A new technique known as the Reflective Online Searching Skills (ROSS) is developed at the School of Information Systems at University of Queensland in Australia. These techniques aim to develop and promote student’s skills and
knowledge of online searching (Partridge et al. 2008, p.57).

2.3.3 Provision of Web-Braille Services

The Web-Braille library service is a breakthrough developed by the National Library Service (NLS) for the Blind and Physically Handicapped at the Library of Congress. The NLS defines ‘Web-Braille’ as follows:

‘(...) is an Internet, web-based service that provides, in an electronic format, many braille books, some music scores, and all braille magazines produced by the National Library Service for the Blind and Physically Handicapped. The service also includes a growing collection of titles transcribed locally for cooperating network libraries. The Web-Braille site is password-protected, and all files are in an electronic form of contracted braille, requiring the use of special equipment for access’ (NLS, 2003).

The project provides more than 2,700 digital Braille books on the Web. For long time, blind library users utilised from the traditional Braille techniques in order to read books. Braille, a system of dots which is read with the blinds’ finger touches, was invented by the French Louis Braille around 1800s. But there is concern that the rapid technological advancements may make the Web-Braille content inaccessible because of intellectual property barriers as will be discussed latter in this Chapter (The Library of Congress, 1999).
2.3.4 DAISY\textsuperscript{27} Digital Talking Books

Another Web-based library service is the DAISY digital talking books (DTB) which is a collection of digital files that provides an accessible representation of the printed book for individuals who are blind, visually impaired, or print-disabled. These files may contain digital audio recordings of human or synthetic speech, marked up text, and a range of machine-readable files. The DAISY DTB utilises the technology of the Web with some specialised applications added to provide greatly improved access to content. There are three types of DAISY, audio, audio and full-text, text and no audio (The DAISY Consortium, 2008).

Unfortunately, provision of such important service in the Sudan is not available for it could contribute to bridge digital divides among the visual handicapped library users.

2.3.5 Cataloguing Web Resources

With the advent of the growth of content published on the Web, an increasing attitude to control access to this content also grows. Bibliographic processes meant to produce a standard catalogue to access any library collections. In the era of the Web, the same tool still functions although it may be manifested in different ways (Paul et al. 2006).

Online Public Access Catalogue (OPAC) replaces the old manual way of content searching and provides more perspectives content. Machine-readable cataloguing (MARC) encodes the information in a computer readable format to facilitate search rather than on traditional

\textsuperscript{27} Digital Accessible Information SYstem.
paper. There have been great efforts trying to create indexes or directories for Websites such as the Librarian Internet Index (http://www.1ii.org). MS-DOS-based library catalogues are replaced by WebLis, Innovative interfaces, imagic DB/Text WebPublisher, Horizon and other Web-based integrated library systems (ILS), (Ibid).

Another advancement in cataloguing is the transformation of worldwide union catalogue, known as WorldCat, within a collaborative strategy between OCLC and libraries of the countries of the world. The strategy has been approved in September 2000 with the aim to transform WorldCat from a bibliographic database into an international networked information resource of text, graphics, sound, and motion. The creation of the WorldCat in Oracle platform establishes a global knowledge base supported by a set of integrated Web-based tools and services that facilitate, contribute to, describe, discover, access, exchange, deliver and preserve knowledge objects as well as the expertise of participating institutions. In order to realise such strategy, OCLC’s primary business units developed new products and services such as metadata, cataloguing, and co-operative discovery techniques (Bellinger, 2002).

In the dawn of the twenty-first century, the Web has added new challenges for libraries in regard to control of its content. The second edition of the Anglo-American Cataloguing Rules (AACR), defines ‘content’ as follows:
'a bibliographic resource that is added to or changed by means of updates that do not remain discrete and are integrated into the whole. Integrating resources can be finite or continuing' (AACR2, 2002).

Integrating resources represent a mode of publishing different from that of monographs and serials. For a long time, most of integrating resources were in print text form as updating publications in loose-leaf format. These resources have been somewhat problematic to catalogue with earlier versions of the AACR2 because such resources exhibit characteristics of both monographs and serials. Prior to 2002 revision of the Rules, cataloguers used the manual 'Cataloguing Rules for the Description of Loose-leaf Publications' by Adele Hallam (Ferris, 2002).

With the dramatic increase in integrating resources available on the Web, it becomes clear that more formal guidance on how to catalogue resources with this mode of publishing is needed. Chapter (12) of the AACR2 has been renamed ‘Continuing Resources’ in the 2002 revision which aims to incorporate rules for serials and integrating resources both finite and continuing (Loc. cit., p.339).

Additions to the MARC 21 Format for Bibliographic Data for integrating resources included a new bibliographic level and specific codes in the fixed fields. Several existing fields in the format formerly used only for serials began to be used in records for integrating resources when the 2002 revision of the Rules has been implemented. These adaptations of AACR2 and MARC 21 are very essential development for controlling the Web-based content to ease access and retrieval. With USD$62,000
grant funding from the US Department of Education, the OCLC has conducted the project for 'Building a Catalogue of Internet Resources' from October 1, 1994, through March 31, 1996 (Library of Congress, 2008).

The project initiates a coordinated effort among libraries and information centres worldwide to create, implement, test and evaluate a searchable database of the USMARC\textsuperscript{28} format bibliographic records, complete with electronic location and access information, for Internet-accessible resources. Voluntarily, libraries participating in the project, in cooperation with representatives from their host institutions, identify, select and catalogue computer files available via the Web and contribute the resulting records to the OCLC. The project tests the suitability and applicability of cataloguing standards and practices, including the MARC communications format and the AACR2 revised edition. All records include one or more 856 fields, 'Electronic Location and Access', which is used to provide access to the described resources, and are suitable for inclusion in local, regional, or national library catalogues (Olson, 2002).

It is noteworthy that the OCLC released the Open WorldCat programme on August 2006 in beta form. It aims to make library resources more visible to the Web users and to increase awareness of libraries as a primary source of reliable information and helpful personal assistance. The programme inserts 'Find in a Library' results within regular search engine results that the Open WorldCat provides a permanent destination page and search box that lets a broader range of people discover the riches of library-held materials catalogued in the WorldCat database (OCLC, 2007, op. cit.).

\textsuperscript{28} US machine-readable cataloguing.
The project has 209 participants from 43 US states, the District of Columbia, and eleven other countries. Some 2,500 records for Web resources are available via the InterCat Catalogue; a Web-searchable, Z39.60-compliant, experimental database built using OCLC SiteSearch and WebZ software. The InterCat Catalogue joins the capabilities of fielded searching with bibliographic description and remote access to Web resources. The database is publicly available at no cost (Loc cit., pp. 3-4).

It is vital to point to one significant assumption that the traditional library-cataloguing unit is simply and irrevocably doomed to extinction and will be replaced by the new model of the Web cataloguing techniques (Bogan, 2003). The trend has shifted towards international cooperative cataloguing in order to expand access to library collections by providing useful, timely and cost effective cataloguing that meets mutually accepted library standards worldwide (Wiggins, 2000, p.18). This is supported by the assumption that digital content just like its physical counterpart should be catalogued and indexed so that users can locate and retrieve them in the future (Cerf, 2005, op. cit., p.13). It is worth to say that the translated Arabic cataloguing rules, which applied in most of the Sudan libraries, still not updated to incorporate rules for cataloguing Internet resources.

2.3.6 The Digital Publishing

One of the greater impacts on library services is the introduction of the e-book. A study (OCLC Report, 2004) reveals that the e-book is the fastest segment of the publishing industry. In the first quarter of 2004, its sales increased 46% from 421,955 units to 288,400 in the first quarter of 2003. Its revenues estimated in 2005
at approximately US$11.8 million among 18 trade and educational publishers, accounting for 5,242 titles (Silberer, 2007).

With smaller number of subscribers to each specialised title, costs have increased from traditionally affordable prices. Commercial publishers are making considerable inroads in the publication of academic journals, subsuming or eliminating many university press activities in this area. Since the late 1980s, costs of subscription to academic journals have increased at an average rate of 13.5% per year. Libraries have been the principal market for academic journals. The acquisitions budgets have not kept pace with such spiralling costs. As a consequence, a crisis in acquisitions of such journals has led to a re-evaluation of the state of scholarly publishing (Thomas, 2002).

The migration of scholarly literature from print to digital formats and the effects this has had on the scientific communication passed, according to Smith (2007), through three stages.

The first stage (1990–1993) followed the introduction of e-journals, both in online and CD-ROM format. At this stage, scientists, authors and users were concerned about quality and sustainability of the new media and many publishers were also hesitant to commit to e-journals.

During the second or evolving stage (1994–2002) the majority of scientific journals became available in digital format, new features were added to some journals, and some individual articles were made available through preprint archives and back issues, and author’s Websites. Most of the e-journals were, however, merely replicas of traditional print journals (some only in digital format, but most published in both formats). Digital preprint services were also generally emerging, together with author self-
publishing initiatives, the development of institutional repositories, and the open archives initiative.

The advanced stage has been reached with the evolution of sophisticated systems that provide advanced capabilities, such as deep links to raw data, individual articles, as well as full-text core journals collections integrated into one comprehensive system.

The above stage may be marked with the growing desire to preserve and document the Web-based content for future access. Though most institutes and universities in the Sudan maintain subscriptions to international e-journal databases but there is no binding preservation responsibilities established yet. There are tremendous threat to databases of e-journal articles and e-books due to many factors such as hacking and piracy that will be discussed further in section (2.4.5).

2.3.7 Digitisation of library materials

Recent advancements in ICTs made it possible for libraries to digitise their print documents and convert them into digital forms. Reasons for digitisation of library materials always attributed to fragility and ephemeral nature of the archival print materials. This process has released content from its media or carrier and is considered one facet of digital preservation. Most of the traditional library catalogues and other services have been migrated to the Web for global reach and access.

In Australia, for instance, the digitisation of a collection of significant newspapers covering a period of 150 years from the early nineteenth century has been completed successfully. Such a project, which would cost around AUD$3 million, would enable researchers to search across the
full-text of all digitised newspapers (Cathro, 2006).

The National Endowment in the US and the JISC in the UK have funded similar projects for the Humanities. A national digital newspaper database could:

- support biographical and historical research, enabling researchers to locate relevant newspaper articles more efficiently than at present, and thus enabling them to focus more on unpublished information sources;
- provide an invaluable resource for longitudinal cultural studies, including media studies; and
- support longitudinal research in certain scientific fields such as ecology and climate change (Minnesota Historical Society, 2007).

Previously, librarians have bemoaned the failure of patrons to realise that not everything is on the Web and that there was life before the Internet. An ongoing digitisation projects, however, are expected to solve the problem. For example, Google, Inc., a company founded by Larry Page and Sergey Brin in 1998, has launched a programme with a number of research libraries in the US and the UK aimed at scanning all the books in their collections (Liedtke, 2006). Further, researches (Lor, 2003) show that the problems of digitisation are not technological only but also economic, political, legal and moral.

On December 14, 2004 Google, Inc. commenced partnership with eminent libraries in the US to make digital copies of their collections accessible on the Web. The project is not new since these libraries already embarked in digitising their materials. Within this decade,
Google, Inc. is trying to digitise more than 15 million volumes to its electronic index at a cost of USD$150 million. Access to these materials is available to researchers as full-text public domain items and bibliographic citations accompanied with URLs of booksellers such as amazon.com or sagepub.com. But Google search display results disabled the users’ save, cut and copy functions in the text to limit full access to copyrighted material (Hanratty, 2005). The same also applies to Amazon.

The Google project has already digitised and indexed many books and made them searchable via Google Book Search. This is known as Google Library Project, formerly Google Print Project that had two facets: the Partner Programme, formerly the Publisher Programme, and the Library Project. Under the Partner Programme, a publisher controls the rights in a book and authorises Google to scan the full text of the book into Google’s search database. In response to the query, the user receives bibliographic information concerning the book as well as a link to relevant text. Such dramatic change in the digital publishing has greater impacts on library acquisitions and collection development hence eliminates book brokers, decreases prices and saves library budgets. The project also provides more options for acquisition librarians to assess, compare, analyse and select better materials for building their library collections properly (Band, 2006).

For the time being, rivals Yahoo and Microsoft MSN\(^{29}\) are working with the Internet Archive, Open Content Alliance and Adobe, Inc. on a competing book digitisation project, which scans public domain works or works legally authorised by their creators (Westin, 2007). Their objective is based on the fact that 50% of online users’ queries go

\(^{29}\) An Internet service launched by Microsoft in August 1995 in the U.S.
unanswered by search engines and that almost 800 megabytes of information is produced per person per year in the US, (OCA, 2005).

The Google project encompasses books in and out of print, in copyright, and in public domain. Participants in the project are libraries of Harvard, Stanford, the University of Michigan, and Oxford University, as well as the New York Public Library (Band, 2006, op. cit.).

Similarly, the British Library and Microsoft Corporation, Inc., have agreed to build the National Digital British Library through digitisation of 25 million pages of content from the Library’s collections about 100,000 of them are in public domain (The British Library’s strategy, 2005-2008, op. cit.).

In the Google Library Project, the New York Public Library contributes its public domain and non-copyrighted material initially to the programme. While Oxford, contributes the nineteenth century collections (public domain) from its Bodleian library to the programme. One of the world’s largest and oldest libraries, the Bodleian has served as a legal deposit for nearly four centuries. Though these efforts are considered giant strides, but participating libraries are worried that publishers might object to the programme on the grounds of copyright violation (Band, 2006, op. cit.). This problem will further be discussed in this Chapter. It is noticed that most of library materials in the academic and public libraries in the Sudan have fallen into public domain since decades though there is no similar cooperative initiatives for digitising them. However, it is noticed that the Google Book Search is more sophisticated and faster in search result display than that of MSN Book Search.

---

30 Available at http://books.google.com/
31 Available at http://search.live.com/books

76
2.3.8 The Rise of cyberlibraries

This part of the study discusses the conceptual foundations for the terms ‘virtual,’ ‘digital’ and ‘electronic libraries.’ The digital library is a new type of library, one that may not necessarily obliterate traditional library activities, but one that certainly creates new services for its clientele. Particularly, digital libraries have the potential to provide an environment in which community members may discover information through means such as knowledge management, data mining, and intelligent agents support dynamic collections, dynamic documents, and collaboration (that is, to allow and foster contribution of materials by community members) be an environment of open structures (Kuny & Cleveland, 1998).

But without doubt, any library business is to connect people with information. An ‘electronic library’ is a library that consists of electronic materials and services. Electronic materials can include all digital materials as well as variety of analogue formats that require electricity to use. A ‘digital library’ is a library consisting of digital materials and services. Digital materials are items that delivered digitally over networks. It is also defined as follows:

‘A distributed information system ensuring reliable storage and effective use of heterogeneous collections of electronic documents (text, graphics, audio, video etc.) via global data transfer networks in a way convenient for the end user’ (Ershova, 1999, p. 78).

The Digital Library Federation (DLF) provides the following definition:
‘Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve, the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities’ (DLF, 2000).

Digital libraries are able to provide for the preservation and migration of electronic information (both digitally born and digitised) as well as enhanced form of library services; i.e., document linking rather than document delivery concept. Both digital and electronic libraries can be virtual libraries if they exist in unreal realm.

An important factor about the impacts of the Web on libraries is the formation of the World Digital Library (WDL). The concept of this initiative came from a speech that Billington\(^{32}\) – the librarian of the Library of Congress – delivered to the newly established US National Commission for the UNESCO on June 6, 2005, at George Town University. Google has agreed to donate three million dollars as the first partner in this public-private initiative (World Digital Library, 2006).

The WDL aims to bring together online rare and unique cultural materials held in the US and Western repositories with those of other great cultures such as those that lie beyond Europe and involve more than one billion people including Chinese East Asia, Indian South Asia and the worlds of Islam stretching from Indonesia through

\(^{32}\) Was nominated in April 1987 by President Ronald Reagan and confirmed by the Senate to be the 13\(^{th}\) Librarian of Congress (Library of Congress Website).
Central and Western Asia to Africa (Lamolinara & Tyler, 2005).

Such initiative will contribute, if managed transparently, to understanding of intercultural relationships and prevent potential clash of civilizations.

2.3.9 Physical v Virtual Library

Terminology in LIS, as with all human constructs, is fraught with ambiguity due to the ICTs impacts. Therefore, it is necessary to clarify and review working definitions for key nomenclature. To discuss coherently the matters at hand, there is a need to agree on the meaning for the following terms: traditional library, virtual library, and digital library.

For the purposes of this study, a library is an entity in which information containers are selected, acquired, organised, disseminated and preserved. Information containers are traditional formats in which information is recorded. To accomplish these tasks, some administrative infrastructure, however intricate, should control the workings of the entity. This definition is precise so that the various purposes of which libraries exist will not be excluded. For example, if we include the educational function that is sometimes associated with public libraries, we may exclude certain corporate libraries whose mission does not include such a function.

At a more fundamental level, any library definition may not be universally accepted. On the one hand, as the virtual library becomes a reality it is apparent that the ‘digital library’ is a complex and dynamic entity (Robins, 2002). The Association of Research Libraries (ARL), (1995) identifies major characteristics of a ‘digital library’ as follows:
• Organisation of digital libraries;
• Collection assessments;
• End-user assessments;
• Network infrastructure considerations for the successful federation of independent databases;
• XML\textsuperscript{33} and related metadata enhancements;
• Interface design considerations;
• An object-oriented interface system; and
• A GIS and patent discipline overviews.

Lancaster (1995, p. 1) points out to three general types of services that traditional libraries provide. Document delivery, question answering, and database searching each can be assessed in terms of effectiveness, cost per successful transaction, risk of incorrect information or a failed search, and benefit of a successful search. Delivery is concerned with direct Web connection of users to information rather than lending items remotely (Oberlander, 2007).

The expectation in a physical library, on the other, is that over time, evaluation procedures stabilise, and knowledge gradually accumulates about what does and what does not work effectively. Lancaster (1995, op. cit.) postulates that though the digital library implies some basic conceptual disparities, it is clearly that document access rather than document delivery which it focuses on. Also the criteria, objectives and assessment of the library use will remain unchanged in essence even if content of these libraries fundamentally differs from those of conventional ones.

At the time Lancaster made the above statement, digital library construction and evaluation were relatively new undertakings.

\textsuperscript{33} Extensible markup language.
One of the remarkable inventions envisioned by the Web capabilities is the creation of the digital library, which made it possible to provide content online remotely. Observably, the concept of digital library is sometimes confused with the automated library approach such as:

'The term ‘automated digital library’ can be used to describe a digital library where all tasks are carried out automatically. Computer programs substitute for the intellectually demanding tasks that are traditionally carried out by skilled professionals. These tasks include selection, cataloguing and indexing, seeking for information, reference services, and so on’ (Arms, 2000).

This confusion is because digital libraries differ from automated libraries of the present day because they depend primarily on digitised or born-digital library materials and operate on LAN portal or intranets or on the Web while the automated library implements computerised catalogues and bibliographic databases but not necessarily connected to a network or the Web. Access to digital library content is ensured mainly via its search user interface which usually served by an intranet or Web portal that contains pane(s) for entering a search expression or syntax, options, drop-down list, Boolean operators for selecting search constraints about such expression or syntax, and sometimes user directions (Ibrahim, 2003). Elliot and Kling define digital libraries as:

‘information systems (IS) and services that provide electronic documents—

---

34 Local Area Network.
files, digital sound, digital video-available on dynamic or archival repositories’ (Elliot & Kling, 1997,p.1023).

In this definition, function of provision of dynamic or archival content constitutes a base and framework for the Sudan Web Archive, which will be planned later in Chapter (4). While Sanchez (1998,p.1) considers digital library as a virtual space in which scholars conduct research, collaborate, and publish their work.

Both definitions conceive of digital library as in some way integrating technology, content, and services, and both imply that complex (that is structured, irreducible, and relatively enduring) ‘things’ are interwoven with peoples’ situated choices and with the available collective resources. The ARL provides that terms such as ‘electronic library’ and ‘virtual library’ are often used synonymously. The elements that have been identified as common to digital library are as follows:

- It is not a single entity;
- It requires technology to link the resources of various born-digital origins.
- The linkages between the many digital libraries and information services are transparent to the end users;
- Universal access to digital libraries and information services is a goal;
- Its collections are not limited to document surrogates, but extend to digital artefacts that cannot be represented or distributed in print formats (ARL, 1995, op. cit.).

However, Bishop (2003,p.297) states that the functions of the digital library are:
• To expedite the systematic development of the means to collect, store, and organise information and knowledge in digital form, and of digital collections;

• To promote the economic and efficient delivery of information to all sectors of society;

• To encourage co-operative efforts which leverage the considerable investment in research resources, computing and communications network;

• To strengthen communication and collaboration between and among the research, business, government, and educational communities;

• To take an international leadership role in the generation and dissemination of knowledge in areas of strategic importance;

• To contribute to the lifelong learning opportunities.

Digital libraries often provide electronic or digital items. The idea of the e-book is first articulated as a ‘memix’ by Vannevar Bush, at the MIT in the seminal 1945 article ‘As We May Think’ (Bush, 1945). Later on, Gutenberg Project began in 1971 to make print material publicly available in plain ASCII format for computers. e-books developed successfully in the 1990s when ‘readers’ software has been introduced for the first time (Dewhurst, 2000).


35 American Sandard Code for Information Interchange.
In 1998, the Rocket e-Book is introduced, and in 1999, Simon, Schuster and King published an e-novella that could be read on any Web browser on virtually any computer, or downloaded to certain e-book services. For the foreseeable future, most electronic publishing may involve scientific, technical, professional, and academic information, as well as some original fiction. Librarians and others involved in digital asset management need to preserve at least some of this material for future reference, since it is expected that original works will be created and many of these may exist only in electronic form. e-books are not a historical object or anomaly, but a new form of content conveyance (Romano, 2002).

It is emphasised that the print conventional library medium is not dead but its obviously that their era of dominance is waning and that the future is digital (Bailey, 2006).

Digital libraries often referred to as 'libraries-without-walls'. As Kuny and Cleveland (1998) criticise, librarians will discover that 'libraries-without-walls' are actually only libraries with new walls, technologically bounded, legally restricted, and administratively hamstrung.

While Mohamed (2006) praises that with digitised services there are many good qualities ought to be appreciated as follows:

- Availability of information is secured, bearing in mind [that] people used to talk previously about the scarcity of information;
- Quick, reliable accessibility, the technology today enabled multiple users to access information simultaneously;

36 First created in 1993 for document exchange independent of the application software, hardware or the operating system.
• Digitisation enables both the user and the indexer to identify and acquire information regardless of its physical location.

These three qualities are regarded as the most important features of library services in the age of information society. Contrarily, the only initiative that can be described as digital library of full-text in the Sudan is the Sudan Open Archive (SUDANAR), which developed on an open source system and mostly contains public domain content by Rift Valley Institute (RVI) with the support from the The United Nations Children’s Fund (UNICEF), and the Sudan Centre for Census, Statistics and Evaluation (SCSE). The Archive is developed on non-legal basis since its content will be withdrawn upon requests of right holders. It can be visited at http://www.sudanarchive.net. Another valuable Web-based content that qualifies as national heritage is the Sudanese Journal of Public Health available at http://www.sjph.net.sd/archive.htm in addition to the publications of the Dams Implementations Unit (DIU) at http://www.merowedam.gov.sd/en/publications.html

2.3.10 Building the Information Society

As referred to by Bowles (1999), the idea of the ‘information society’ has emerged as a result of the exponential increase in the quantity of information seems to be a phenomenon similar to that of the ‘information explosion’ after the Second World War.

But Masuda (1981) indicates that one of the most interesting actions have occurred in Japan, where in 1972 a non-profit organisation called ‘the Japan Computer Usage Development Institute’ presented to the government ‘The Plan for
Information Society - A national goal toward the year 2000’. This plan had been developed for presentation as a model plan for the realisation of Japan’s information society. It gave a picture of an information society that was desirable and could be realised by 1985. The goal of the plan, Masuda goes on, has been the realisation of a society that brings about a general flourishing state of human intellectual creativity, instead of affluent material consumption marked by the dominance of the capitalism culture. The Plan includes an ‘Intermediate Impact Plan’ that has required an investment of three million dollars between 1972 and 1977, and a ‘Long Term Basic Plan’ that calls for an investment of USD$65 billion between 1972 and 1985.

Since then many new terminologies have emerged in parallel with the term ‘information society’, inter alia, the ‘post-fordism’, ‘post-industrial society’, ‘postmodernism’, and the ‘telematic’ societies. Many theorists have been studying and focusing on such concepts and their implications such as Manuel Castells, Daniel Bell, Frank Webster, Anthony Giddens and others (Ibrahim, 2005, op. cit.).

However, due to supremacy of the US in the administration of the Internet domain names, global security and safety concerns, and the digital divide phenomena, have led the United Nations to organise the World Summit on Information Society (WSIS) in two phases in Geneva 2003 and Tunisia in 2005 in order to address such challenges in which libraries play core roles in connecting people with information on global basis (Ibrahim, 2005, op. cit.).

The concept of ‘information or knowledge society’ is strongly opposed by many studies for being a new propaganda aiming to impose principles of the capitalism in the guise of globalisation. The self-styled ‘information
societies’ or ‘knowledge societies’ are in fact neologisms which hide the ideology of the dominant classes of capitalism in its most violent imperialists stage (Muela-Meza, 2006).

2.3.11 Web-based Content as Cultural Heritage

Increasingly, Web-based content is recognised as cultural heritage and enshrined in Article (1) of the UNESCO Charter on the ‘Preservation of the Digital Heritage’ as follows:

‘The digital heritage consists of unique resources of human knowledge and expression. It embraces cultural, educational, scientific and administrative resources as well as technical, legal, medical and other kinds of information created digitally, or converted into digital from existing analogue resources. Where resources are ‘born-digital’ there is no other format but the digital object. Digital materials include texts, databases, still and moving images, audio, graphics, software and web pages, among a wide and growing range of formats. They are frequently ephemeral, and require purposeful production, maintenance and management to be retained. Many of these resources have lasting value and significance, and therefore constitute a heritage that should be protected and preserved for current and future generations. This ever-growing heritage may exist in any language, in any part of the world, and in any area of human knowledge or expression’ (UNESCO, 2003, pp.61-62).
The first component of Web-based content is the 'Web site', which is understood to be a collection of Web pages. OCLC provides two different definitions for the term 'Web site'. One of these definitions rely strictly on physical (network infrastructure) criteria, and the other on information (content-oriented) criteria:

'Web Site (Physical Definition): the set of Web pages located at one Internet Provider address.
Web Site (Information Definition): a set of related Web pages that, in the aggregate, form a composite object of international relevance' (OCLC, 1999).

The Web site may contain surface content or deep Web content such as databases of open or restricted content. In the extreme, the ‘linked’ nature of the Web opens the door for an argument that the Web itself is one grand Web site.

According to a UN Report (2004), the number of the Websites around the world has reached 51,635,284 million on June 2004 with 26.13% compared to 2003. While the number of sites use the Secure Socket Layer (SSL) protocol increased to 56.7% between April 2003 and April 2004 to reach 300,000 Web sites. Petrov (2004), states that Web content or Web-based content simply is Information, document, and data published over Internet technology.

Since the invention of the Web there are no serious studies on Web-based content aim to provide statistics, analysis, description and coverage. The OCLC has launched an annual project known as 'Web Characterization Project' in order to address questions of general interest, such as 'how big is the Web?' and 'what kinds of
information are available on the Web? and more in-depth, library-oriented issues, such as the rate of migration of print materials to Web-accessible formats and the dynamic properties of Web resources (Lavoie, 1999; Bjorneborn & Ingwersen, 2001, op. cit., Noruzi, 2006). Unfortunately, the project did not continue since 2002.

The majority (72%) of the content on the Web is in English language, 7% is in German, 6% is in Japanese, 3% is in both Spanish and French, 2% is in Italian, Dutch and Chinese, and only 1% is in Korean, Portuguese, Russian and Polish respectively (O’Neill et al. 2003).

User-created content (UCC) is a new social impact identified as content made publicly available over the Web (e.g. wikis\(^{37}\), text blogs, Web 2.0\(^{38}\), videos, podcasting\(^{39}\), games, films and music), which reflects a certain amounts of creative effort and is created outside of professional routines and practices. Most of UCC activities are undertaken with no expectation of remuneration or profit. Their motives include connecting with peers, self-expression, and achieving a certain level of fame, notoriety or prestige. There are five basic models of UCC content:

---

\(^{37}\) A Website that allows users to add, remove and otherwise edit and change content. An example of this is the online encyclopedia Wikipedia that comprised 4.6 million articles in over 200 languages in 2006.

\(^{38}\) A new application defined as ‘principles and practices that are founded on services rather than software, user control and user participation, scalability from small to large, remixable data sources and data transformations, and the harnessing collective intelligence. These practices include file sharing and social networking such as YouTube, Facebook, blogs, RSS, wikis and others (O’Reilly, 2005). Campbell (2007) argues that many of these tools have analogues in the library community. File sharing with MARC is used since 1960s as well as other tools.

\(^{39}\) A combination of ease of audio production with technologies that allow for subscription and syndication.
• Voluntary contributions;
• Charging viewers for services, e.g. pay-per-item or subscription models, including bundling with existing subscriptions;
• Advertising-based models;
• Licencing of content and technology to third parties; and
• Selling goods and services to the community.

This type of content could enrich political debate and participation, education, human rights, transparency and cultural and social development. But it raises challenges of infringement of intellectual property rights, severe exclusion (between elderly, handicapped, illiterate and poor people), cultural fragmentation, quality (the fact that most authors of this content do not cite sources of references for documentation and authenticity or there is no editorial and peer-reviewing standards), privacy, safety and security in a participative Web of social networking sphere (OCED, 2007, pp.9-12).

It is observed that presence of Web-based content of the region, in which the Sudan belongs to, is less visible in proportion to global Web community. Noruzi (2006) attributes such phenomena to the notion that most Websites of this region lack links and indexing techniques often applied by search engines.

Nonetheless, Web-based content has become a new responsibility of the national libraries to collect, process, preserve and make it accessible. Usually this content resides in two layers surface and deep. Sometimes these layers are called ‘surface Web content’ and ‘deep Web or invisible Web content’ and they fall into the domain of library materials.
2.3.12 Deep Web-based Content

The deep Web contains pages that are transmitted by Web servers in response to a client’s queries. Both the terms ‘Invisible Web’ and ‘Deep Web’ have been coined for the vast quantity of Web pages that are accessible with a Web browser but are difficult to find because they are of a type not indexed by search engines (Bergman, 2000).

Deep Web-based content often reside on restricted Web databases but not retrievable using search engines that rely on crawlers or spiders, for example, data in file format such as PDF, database content accessible only by query, and information contained in frames. The number of documents available in the deep Web is estimated to be 400-500 times greater than the amount of content retrievable via conventional search engines (the ‘surface Web’), with over half of the ‘hidden’ content residing in topic-specific searchable databases. The deep Web content term is also used for password-protected Web-based content available only to authorised users, i.e., members, subscribers (Reitz, 2004).

Search engines crawlers, for example, are adept at finding Web pages, images, PDF files, and other types of static content on the Web. This type of document has fixed address or URL, and unless it’s moved, the URL will always point to the exact location of the content on the Web, making it a trivial task for Google to fetch the page.

By contrast, dynamic content poses a very difficult challenge for the search engines. It is stored in databases and has no fixed address or URL and because it lacks such address, and without links, crawlers cannot locate content (Sherman, 2005).
At a time of ever increasing demand for information and its ready availability, libraries need to keep ahead of developments in Web technologies and advanced ICTs systems to provide the services that today’s society both needs and expects. Libraries can no longer afford to be parochial and isolated. They should cooperate to share resources since the Web imposes per se a new dimension and a new legal obligation for preserving cultural heritage for future generations.

2.3.13 Content Industry

Digital content industry is emerging as a new economic sector with tremendous expertise and creative potential for the least developing countries. This new sector has been brought about through the convergence of previously distinct areas such as traditional library content, media and entertainment, software, hardware, multimedia, and telecommunications. Digital content encompasses the creation, design, management and distribution of digital products and services and the technologies that underpin them.

The Australian Digital Content Industry Agenda (ADCIA), looks at ‘digital content industry’ as the production and marketing of cinema and broadcasting programmes in the form of digital and interactive television, online games, reusable electronic educational content, museum and galleries artefacts, libraries and archives content, Web-based products such as music, text, and films, in addition to the development of software, and online services that aim to create and publish digital media and visual effects (ADCIA, 2005).

The digital content industry, while still at relatively early stage of development, is
emerging as an area of significant global opportunity valued at more than USD$172 billion in 2001 and estimated at more than USD$434 billion in 2006, an average annual growth rate of almost 29% (FORFAS, 2002).

The global growth and size of the media and entertainment sectors have been forecast to grow at an average annual rate of 7.3%, from USD$1.8 trillion in 2005 to more than USD$2.4 trillion by 2009. Many of these sectors are driven by digital content such as the Internet, games, and business information. The Australian digital content industry, for example, is estimated to produce output worth AUD$21 billion, almost 3.5% of Australia’s GDP\(^40\) compared with the UK and US, where GDP shares are conservatively estimated at 5% and 7.8% respectively, and employs about 300,000 persons. Australia plans to achieve a sustainable and internationally competitive digital content industry which doubles in value to AUD$42 billion by 2015 (Strategic Industry Leaders Group, 2005, op. cit., p.8).

A Report on ‘Arab Human Development’ discloses that the Arab countries face series problem in production and distribution of content in Arabic. It attributes the causes of the problem to the high illiteracy rates and lack of funds, support and encouragement of research and development (R&D). Also attributed to the problem of digital divide represented in poverty, shortages of expenditure on R&D, brain drain and other factors. For instance, average expenditure on R&D in Arab countries collectively does not exceed 0.2% of the GNP\(^41\) (UNDP & RBAS, 2003).

Studies show that Internet penetration in the Middle East and North Africa is only 10% of the total population. The reason for this low penetration rate is due to lack of content and

\(^{40}\) Gross Domestic Product.

\(^{41}\) Gross National Product.
Arabic search engines. It is estimated that there are around 300 million pages in Arabic compared to 30 billion pages available in English (Dunworth, 2008). Estimates also reveal that Arabic language content account for only 0.3% of the total content available on the Web (Abdulla, 2008, p.5).

The previous Arab Human Development Report also enumerates three categories of Arab institutions that focus on R&D. The first are higher education institutions and their affiliated research centres, the second are free-standing specialised centres of scientific research, and the third are the research units with links to industry, all summed up at 588 throughout the region. There are 184 Arab universities, all involved in promotion of R&D. It is concluded that the defect in Arab scientific R&D agencies is their inability to transform research results into investment tangible projects. This defect is linked directly to media institutions, which also produce content, for most Arab media agencies do not have reliable information centres housing libraries and archives (UNDP & RBAS, 2003, op. cit., p.60).

The situation in the Sudan does not differ mostly from the Arab world. Though recently a Ministry of Technology has been established to carry out the mission of advancement of R&D in the country, but still content industry is in its infancy. The new Ministry is composed of the previously fragmented research centres though there is little awareness of the importance of establishing an agency to maintain R&D researches on indigenous knowledge content.

However, the recommendations of the Workshop on Internet Issues organised by the National Telecommunication Corporation (NTC) encourages production of local content in the Sudan as follows:
• Encouraging institutions and individuals to form, build, host and register local Websites;
• Urging all government authorities and companies to provide their services through the Internet;
• Encouraging and supporting all initiatives aiming to produce local content through the Internet;
• Enacting suitable laws to provide access to the information of the government provided that this information do not affect the national security and law;
• Encouraging the increasing of content in Arabic language and local dialects;
• Activating the intellectual property laws to protect local content (NTC, 2007).

Inferred from the Arab Human Development Report (UNDP & RBAS, 2003, op. cit.) as mentioned earlier, content industry in its broadest meaning deals with three essential levels:

• production of innovations, i.e. electronic hardware, telecommunication devices, pharmaceutical industries and others.
• production of intellectual works such as essays, articles, drama, music, media programmes and entertainment.
• applying tools and mechanisms of data or information engineering such as producing bibliographies, abstracts, catalogues, and indexes which control flow and access to content.

More importantly, online content is increasingly being recognised as a source of historical, educational and cultural heritage and as such,
libraries and other information services are beginning to move towards becoming ‘digital libraries’ (FORFAS, 2002, op. cit.).

Unfortunately, the practical notion for successfully managing digital libraries is not yet mature in the Sudan. This is clearly manifested in the ongoing local projects that aim merely to set up Internet cafés as e-libraries or bibliographic databases (as in most academic libraries) based on an open source systems as an end in itself. In this sense, two main hurdles are faced: first open source systems face interoperability failures over the Web environment, second if print sources being digitised to migrate the full-text work then an intricate problem of legal rights will frustrate the mission.

To promote production of content, the Library of Congress (2008) set up a strategy towards 2013 expecting to have the following in place:

- Increased digital content holdings;
- Capability to produce and receive sustainable digital content from multiple sources;
- A stewardship network of collaborative partners;
- Recommendations on digital content information architecture, preservation and access;
- Recommendations on public policy for digital content preservation and access.

Also the Committee on Coordination of the Electronic Library Project (CCELP) in the Sudan emphasises that there are three fundamental technical bases that should be maintained for enabling growth of content industry namely:
• the bibliographic control; how to control the tremendous flow of content through capturing, collecting, and identification;
• indexing, i.e. keyword control for recalling content of the document in order to meet the user’s needs;
• subject analysis in order to specify and link the subjects in certain document 42.

There is another type of content, which needs to be considered when planning for promoting content industry known as indigenous or traditional knowledge (TK) and folklore. Hansen and VanFleet define the term ‘traditional knowledge’ as follows:

‘(...the information that people in a given community, based on experience and adaptation to a local culture and environment, have developed over time, and continue to develop. This knowledge is used to sustain the community and its culture and to maintain the genetic resources necessary for the continued survival of the community’ (Hansen & VanFleet, 2003, p.1).

Also the term is defined as follows:

‘(...knowledge passed on from one generation to another usually orally. The oral nature and intergenerational cross-fertilization of traditional knowledge contributes to its value’ (Kawooya, 2006, p.6).

Exploitation of the TK products has raised dispute of legal rights ascribed to its creators as will be
discussed in section (2.4). It is evident that preservation of Web-based content could contribute to production of more content and assists in flourishing content industry sector. But the impacts of the Web on library and information services get complicated when issues of legal rights arise as will be discussed in the next section.
2.4 Barriers of Intellectual Property

The investigated researches on the second dependent variable (see Chapter (1), section 1.9.4) show that legality of copying for preservation purposes constitute major challenge. Publishers, authors and inventors often seek economic and moral rewards when they produce their works and distribute them for sale on the Web. These works often introduced or labelled with whole, part or combination of symbol identifiers and notices similar to '©All rights reserved', or 'No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior permission of the publisher'. Or 'Registered patent®', or 'Registered trademark™', or 'Registered name' or 'Service markSM'.

These identifiers and notices have been a matter of controversy since recent advanced technologies have made it easier to reproduce, store and redistribute any documentary heritage or content over the Web. Such developments made publishers and other rightholders to encrypt their content and apply certain technological measures that block access to it. As a result, exceptions and limitations for copying for national library preservation programmes are at stake.

Documentary heritage includes a wide variety of published and unpublished material protected by intellectual property rights (IPRs). In order for the national library to preserve these material, it is often necessary to copy it. To avoid infringement of these rights, the library can comply with special conditions under such rights, copy only material that are in the public domain,

43 Both TM and SM symbols represent unregistered trademark or service mark, while © symbol is a warning notice to advise that the mark is registered and their use provides legal benefits. A trademark is often referred to as a brand (INTA Website).
or track down the copyright owner and seek permission or pay a fee. Each of these avenues are considered unfeasible. The normal way, as this study advocates, is to be reinforced by a Web legal deposit jurisdiction (cf. Chapter (4), section 4.4.1).

The IPRs including copyright have significant effects on the nature and extent of information services national library provides to their users. Worldwide, those laws are currently undergoing major revisions in response to the growth in the use of digital content in e-governance, e-commerce, e-education and other e-activities. Principal objective of this section is to analyse and interpret the legal impacts on digital preservation.

IPRs law is a complex topic and it is not the intention of this study to provide full coverage to. Indeed, it assumes that the reader understands some copyright basics, such as the notion that facts and ideas are not protected by copyright. Rather, the aim is to highlight some key factors about how and why the IPRs and related rights have evolved and how this constitutes problems to preservation of Web-based content at hand in comparison to the legal deposit principles. Though this study is to focus on preservation of Web-based content in the Sudan National Library but the issues being investigated are of global concern and initiatives in other countries may provide guidance and solution if co-operation is maintained. The most important function of the national library is to preserve the country’s heritage in any format.

In the twenty-first century, the paramount role of libraries should be as Kim (2006), the former President of the Republic of South Korea stresses, to provide economically people with special needs and nations with opportunities for intellectual contacts at the highest level. No other
institutions can take this crucial role in removing dark shadows of the ‘information divide’.

Some studies (Jones, 2003) note that legal implications of Web archiving are key concern of those wishing to preserve Web-based material. The most influential implications posed by IPRs are those of TRIPS and GATS (they will be discussed herein later as part of the problem as stated in Chapter (1), section (1.1)). This is because such content include highly proprietary databases or material such as e-books, scientific and professional e-journals, e-patents, e-industrial designs, software, domain names, legal cases, multimedia and entertainment and other useful items. Preservation of these items could enable transfer of technology from generation to another hence promote economic growth and sustainable development.

Bailey (2006, pp.116-117) argues that the Internet functions at its most fundamental level as ‘content-application, and hardware-neutral network’. He goes on saying that in the recent years, network neutrality has come under threat of restrictions of laws. In other words, the Internet is a neutral open platform this is because it does not prefer certain data packets or deny certain access to packets. The IPRs barriers are discussed as follows:

2.4.1 Challenges of Access to Information

In order to understand the impacts of the IPRs on digital preservation in general and Web-based content in particular, it is important to shade light on the term ‘intellectual property’. The WIPO defines it as follows:

‘creations of the mind and is divided into two categories: ‘industrial

44 Also called ‘intangible property’. (INTA Website, op. cit.).
property, which includes inventions (patents), trademarks, industrial designs, and geographic indications of source; and copyright, which includes literary and artistic works such as novels, poems, plays and computer programs, films, musical works, artistic, artistic works such as drawings, paintings, photographs and sculptures, and architectural designs. Rights related to copyright include those of performing artists in their performances; producers of phonograms in their recordings; and those of broadcasters in their radio and television programs’ (WIPO, n.d.).

The researcher notices that in the WIPO’s above definition there is no reference to creations of the mind published on the Web. This has adverse effects on amendments of national IPRs legislation and may not assist in preservation of the digital heritage before it fades away. Based on the definition, it is useful to differentiate between two categories: the ‘industrial property’ and ‘copyright,’ both of them will be discussed hereinafter for they relate directly to the preservation programmes.

The international scale of protection of IPRs date back to international agreements and conventions such as Paris Convention (1883) on Industrial Property and the Berne Convention on Copyrights (1886), while recent international agreements include the Rome Convention (1961) on Recording of Performing Arts, the Budapest Treaty (1977) on Deposit of New Micro-organisms for the purpose of Patent Procedures and the Washington Treaty (1989) on Integrated Circuit Layout Designs (see also Appendix B).
Writers have devoted a considerable effort to criticise the theme of IPRs. Kuruk (2002), among others, views the IPRs as one of the colonial aspects. This is due to the fact that Africa’s historical communal or collectivist approach to ownership of creative expressions associated with oral tradition is conceptually different from the ‘Western’ orientation that ascribes ownership to individuals and that indigenous African communities look at property as heritage which should be transferred to the coming generations as an ideal method of innovation.

Based on the researcher’s understanding, it is clear that administration of IPRs by two different international organisations, as outlined latter on in this section, might causes some legal confusion and leaves grey room for dispute concerning prospective responsibility of managing IPRs on the Web. For instance, WIPO is a UN agency for administration and enforcement of IPRs, while accession to the WTO, another UN agency for administration of international bilateral trade protocols, requires that the same member states of WIPO would be expected to become parties to the TRIPS agreement and take on the obligations applicable to other member states at their respective level of development.

TRIPS agreement, in turn as Abbott and Correa (2007) point out, obliges members to comply with both Paris and Berne conventions with few exceptions though least developing countries have been granted special and differential treatment pursuant to Article (65.5) and (66.1) of the agreement. The application of advanced digital or technological protection measures (DPM or TPM) to copyrighted works prohibits provision,

45 DPM or TPM is ‘a collective name for technologies that prevent library patrons from accessing a copyrighted digital work beyond the degree to which the copyright owner (or a publisher who may not actually hold a copyright) wishes to allow them to use it’ (Bailey, 2006, op. cit., p. 117); e.g. VitalSource Bookshelf applied by Elsevier.
preservation and access to digital content in general. They limit access to derivative works, fair use and other statutory exceptions and limitations (OECD, 2007, op. cit., p. 88).

The traditional areas of patents, industrial designs, and trademarks, have been joined by various other types of intellectual property such as trade secrets, utility models, broadcasting rights, and rights on compilations of data. This great expansion is strongly reinforced by industrialised countries. It is premised on the idea that their future economic success depends basically upon their superior new knowledge. Recent technological developments, particularly in biotechnology and ICTs, proved that knowledge is principal force of competitive advantage for both private sector and the public sector (Musungu & Dutfield, 2003).

Though its content is increasingly exhausting, the current situation of indigenous or traditional knowledge (as referred to earlier as TK) is becoming the object of growing interest by scientists. TK owners complain about the misappropriation of their knowledge and demand two sets of remedies. First, they ask for some protection of their IPRs. Second, they demand the respect of their customary laws, arguing that they do not ask to be granted new rights but to have their existing rights respected (Brahy, 2006). An emerging concern is raised, as discussed in section (2.3.13) regarding protection and preservation of TK on the Web.

In Africa, as the researcher points up, TK might transcend geographical borders of one ethnic group into neighbouring indigenous communities. Such assumption is evidently vastly applicable to the Sudan though most of it awaits to take its way to the Web to contribute to the overall national heritage.
Earlier in 1970 a number of African countries have adopted the protection of folklore as part of their IPRs. International efforts concerning protection of this content continued until the historical meeting of the World Forum on the Protection of Folklore, held by WIPO and UNESCO in 1997 (Kuruk, 2002, op. cit.).

Recently, there are a number of initiatives at regional and national levels to provide legal protection to TK and associated innovations. Researches such as those of Mugabe et al. (2001) criticise the current forms of IPRs protection for being established to cover industrial innovations. Their structures and requirements are inimical to the protection of TK and expressions of folklore. They are also not suitable for the regulation of access to genetic resources, and the promotion of benefit sharing as envisaged by the Convention on Biological Diversity (CBD), December 29, 1993.

The way that the TRIPS agreement has internationalised the IPRs protection reveals that technology owners protect their works in all areas of technology. Consequently, there are adverse implications of such protection in the context of human rights to food, healthcare and access to information. To some extent, African countries suffered from ‘biopiracy’, a term which is defined as:

‘(...) the concept which covers the dichotomy between the absence of protection for traditional knowledge in IPRs frameworks and the protection afforded, for instance, through patents to innovations that are derived from traditional knowledge in a western scientific mode’ (Cullet, 2005,p.7).
Some scientists who contribute to availability of TK protect their works under IPRs and claim ownership of the resultant artefacts given the intellectual effort vested in creating new knowledge from the existing TK (Kawooya, 2006, op. cit., p.10).

As a result, holders of the African TK content seem pessimistic about IPRs protection of TK because indigenous expressions of knowledge and culture have often been misappropriated based on the argument that they were in the public domain (Armstrong & Ford, 2005).

Since the 1970s, misappropriation of indigenous content led the UNESCO and WIPO to establish agendas for preservation of cultural heritage of underprivileged societies. The UNESCO Convention on the Protection and Promotion of the Diversity of Cultural Expression adopted on 20th October 2005 is one of the most recent initiatives on TK content. The Convention identifies the place for cultural artefacts in the market place and the role of a country in controlling aspects of their cultural heritage from misappropriation and misuse through market mechanisms (UNESCO, 2005).

Following the UNESCO and WIPO’s proposal of ‘sui generis’46 system of protection tailored to the communal natural of TK content, unbounded by time, predominately oral in character and strong linkage to traditions and norms, there is renewed interest in that approach for protection of such content in Africa (Armstrong & Ford 2005, op. cit.; Britz & Lipinski 2001; Githaiga 1998).

Mugabe et al. (2001, op. cit.), contends that the IPRs protection carves out exclusive rights to an individual (either individual natural person or a legal one) to exploit particular creations of ingenuity.

46 The phrase ‘sui generis’ stems from Latin ‘sui’ which means ‘of its own’ and ‘generis’ the genitive form of ‘genus’, meaning ‘kind’. Sui generis then means ‘the only example of its kind; unique’, (Estep, 2003, p.20).
The ‘sui generis’ system should align with the copyright law if benefits of the preservation of scholarly TK-related content are to be fully realised. Some researches (Brahy, 2006, op. cit.) emphasise that there are strong arguments in favour of the creation of a database for preservation of African indigenous TK content. After this review, the next section describes the impacts of the industrial property rights.

2.4.2 The Industrial Property Rights

As discussed above, the industrial property includes essential subcategories for identifying or protecting products, goods and services from piracy and fraud. The most important categories of IPRs that directly concern preservation of Web-based content are the protection of patentable protocols, databases and hardware applications. Dixon and McCowen define the patent as follows:

'A patent is a privilege granted by a government, allowing the holder to exclude others from making, using, importing and selling an invention. Patents provide the holder with an effective monopoly on a particular product or production process. These privileges apply in the countries where they are granted for a limited period, a minimum of twenty years’ (Dixon & McCowen, 2006, p.1).

Patents are the source of serious costs to developing countries but considered a source of wealth for industrial countries. According to study (Dace, 1999), two-third of the estimated USD$7 trillion market value of all publicly traded American companies lies not in their real
estate or plants but in their patents and other intellectual assets. Michalopouls (2005), however, accused the developing countries of stealing most of the patents in the developed countries (though the reverse is also the norm). The researcher assumes that if a product requires considerable ingenuity as well as investment in time, money and effort to produce, but can be copied easily, then there may be little incentive to invent, and too few inventions from the viewpoint of the public interest. As an inseparable part of IPRs, the next paragraphs investigate the protection of the domain names as they directly relate to Web archiving.

2.4.2.1 Protection of Domain Names

While designed to serve the function of enabling users to locate servers in an easy manner, DNs have acquired further significance as business identifiers and, as such, have come into conflict with the system of business identifiers that existed before the arrival of the Internet and that are protected by IPRs (WIPO, 1999, op. cit.).

One thread in the fabric of discussions and consultations concerning management of the DNs has been the interface between the DNs as addresses on the Internet and the IPRs or, more specifically, trademarks and other recognised rights of identity as they had existed in the world before the arrival of the Internet. As discussed in section (2.2), the DNs and the gTLDs are brought into conflict with trademark law. In 2001, legal disputes reached 3,260 cases from 56 countries (Wilbers, 2001). The WIPO (2004) reports that the ‘.com’ domain is by far predominant on the Internet at 84% in 2004.
It has become apparent to all that a considerable amount of tension has unwittingly been created between, on the one hand, addresses on the Internet in a human-friendly form which carry the power of connotation and identification and, on the other hand, the recognised rights of identification in the real world, consisting of trademarks and other rights of business identification, the developing field of personality rights, whether attaching to real or fictional characters, and geographical indications. One system, the DNs, is largely privately administered and gives rise to registrations that result in a global presence, accessible from anywhere in the world. The other system, IPRs regime, is publicly administered on a territorial basis and gives rise to rights that are exercisable only within the territory concerned (Albert et al. 1999, op. cit.). The coming part shades light on the nature of the copyright regime.

2.4.3 The Copyright Law

This law has emerged as one of the most important means of regulating the international flow of ideas and knowledge-based products. It is intended to encourage the creation and dissemination of author’s work to promote cultural and economic development (OECD, 2007, op. cit., p.77).

The philosophy behind this law is to provide protection for literary and artistic works, giving their creators the ability to control certain uses of their works. The law of related rights (or neighbouring rights that is, rights related to copyright) provides similar protection for the creative contributions of parties involved in presenting works to the public, such as performers, phonogram producers, and broadcasters (WIPO, 1999).
Copyright and related rights often established by national laws in individual countries. International treaties link the various national laws by ensuring that at least a minimum level of rights will be granted to creators under each national law. The treaties do not themselves grant rights, but rather require the countries that join the treaties to grant certain rights specified on a non-discriminatory basis. But what is copyright?

IFLA Committee on Copyright and other Legal Matters (IFLA-CCLM) defines ‘copyright’ as follows:

‘a person’s exclusive right to authorise certain acts (such as reproduction, publication, public performance, adaptation etc.) in relation to his or her original work of authorship. The creator of the work typically owns the copyright, at least initially’ (IFLA-CCLM, 2004,p.2).

Another simple, short and precise definition of ‘copyright’ is ‘the right to make copies’. (Loffman, 2003,p.2). Ironically, Nicholson (2007,p.2) views copyright as ‘a statutory monopoly and is characterised by its international nature’. The author of a work, therefore, initially owns the right and, in the case of a published work he transfers some or all of his rights under copyright to the publisher to make certain amounts of copies. As a consequence, the publisher then becomes another rightholder. Only the right holder can authorise the making of copies although his publisher or his agent often do this on his behalf.

But what happens if the work becomes out of print? As Quint (2005) elaborates, in the event that the publisher concedes that the work is out
of print and he does not intend to reissue it again, all rights granted to the publisher by the author shall revert to him pursuant to notification in writing by the publisher.

2.4.3.1 Impacts of Copyright Law

There is asymmetrical relationship between categories of IPRs and the materials covered by the legal deposit laws or national library acts. The role of the national library is to preserve and provide access to the same content of copyright categories as evidence after obtaining legal exemptions in accordance with the amendment of national copyright legislation.

Apparently, the global scale of Web publishing does raise doubts as to the possibility of implementation of the copyright laws of each country on the Web. The relationship between copying for national library preservation purposes and the copyright protection on the Web is governed by an intricate legal nature. Lucas (2005, op. cit.), for instance, illustrates this when he states that the issue of conflict of laws, in the field of copyright, has long been neglected. He goes on that it might be also thought that the conflict of laws could be settled through the harmonisation of legislation and through the standard principle of international copyright conventions, namely the principle of assimilation of foreign works to nationals or 'national treatment'. Pursuant to this principle, it is legitimate that wherever an author resides in a foreign country he is allowed to enforce his or her own copyright under that country’s domestic copyright law.

On the cyberspace, there is a challenge concerning preservation and access to content
through remote lending, copying and document delivery services as enshrined in the legislation of print national library material. For example, Section (109) of the US Digital Millennium Copyright Act of 1998 prohibits circumvention and imposes criminal penalties. Accordingly, users who enjoy lending rights with respect to works that were subject to the first sale doctrine\textsuperscript{47} because they have been purchased outright, currently face licencing and copyright barriers. The TPMs or DRMs measures applied to prevent unauthorised access may allow copyright holders to control use and disposition of digital works long after the protection duration is over and this evaporates the right to public domain (Gaubiac, 2002). As a result, access through licencing to national library collections affects the information services as discussed previously.

Increasingly, copyright protection stifles creativity and innovation and drastically undermines the opportunities for learning and knowledge diffusion to the least developed countries that industrialised counterparts had previously enjoyed (Musungu, 2005).

Over the time, implementation of the above-mentioned measures for software encryption as a copyright protection requirement might become obsolete due to technological rapid change. Therefore, Web-Braille content, for instance, may go missing forever. This access barrier is not being considered while negotiations of

\textsuperscript{47} A legal principle that limits a rightholder's rights to control content after it has been sold for the first time. According to first sale doctrine, lawful ownership of an item, such as music CD or a book is not the same as owning the copyright of the item. The owner of the item may lend, resell, give away and or destroy the copyrighted item but does not grant the right to copy the item in its entirety. The transfer of the copy does not include the transfer of the content's copyright. The legal principle applies to physical items as well as digital content that is downloaded over the Internet (Webopedia Website).
accession to WIPO and WTO are submitted by member states (Roos, 2004, op. cit.).

It is argued that most African library preservation projects do not have a specific copyright policy as such. IPRs protection of TK and other institutional resources, for instance, remains a grey area in Africa. This is because the copyright regime is embedded in the Western individualist ownership system. But TK, which much of Africa’s scholarship intersects with, may not be protected because it fails certain copyrightable criteria including proven ownership (Gladney, 2007, op. cit., p.18).

It is claimed that current copyright regimes are unable to fully accommodate and protect indigenous folklore worldwide and this attitude supports Gladney’s arguments as discussed above. This, as Githaiga (1998, op. cit.) claims, is due to the defects of the law in dealing with requirements appertaining to ownership and authorship, material form, originality, duration, and rights in derivative works.

Interestingly, Pedley (2003, p.47) made an observation that librarians and information specialists find themselves in a critical situation playing the role of ‘piggy in the middle’, working to protect the intellectual property and in the same time are committed to communicate information and new ideas to users of library and archives.

Another new approach to copyright is the ‘Creative Common’ licences such as ‘attributive’, ‘non-commercial’, ‘non-derivative’, or ‘sharing’ (Perry & Callan, 2006). This approach is recently added to

---

48 Someone who is involved in an argument between two opposing sides, but does not know who they should support (Longmann Dictionary of Contemporary English, (online)).
terminology of IPRs on the Web that provides licencing to materials published and accessed online. But the copyright law provides some unrestricted areas as will be discussed later.

2.4.3.2 Limitations and Exceptions

The main objective of limitations and exceptions in IPRs legislation is to ensure access to certain groups of communities due to exceptional reasons known as ‘fair use’\(^{49}\) or ‘fair dealing’.

Exemption from copyright liability provides national libraries and archives an explicit certain provisions not to be infringed. Generally, non-infringing copying for non-commercial purposes should comply with the following:

- be a single copy;
- made by a library or archive or by staff of such acting within the scope of their employment;
- not be associated with any commercial purpose;
- be copied from a collection that is open to the public or at least to all users; and
- include a notice of copyright (Hanratty, 2005, \textit{op. cit.}).

The most significant and, perhaps, obscure limitations on a copyright owner’s exclusive rights is the doctrine of ‘fair use’. This doctrine is an affirmative defence to an action for copyright infringement. It is potentially

\(^{49}\) The US Congress adopted the first fair use statute in 1976. It aims to grant certain privileges restricted to purpose, nature, amount and effect of use for given copyrighted work (Crews, 2006, p.44). It is known as ‘fair dealing’ in the UK.
available with respect to all manners of unauthorised use of all types of works in all media. When it exists, the user is not required to seek permission from the copyright owner or to pay a licence’s fee for use. Section (108) of the 1976 US Copyright Act as amended, for instance, provides that in certain circumstances and under certain conditions it is not an infringement of copyright for a library or archives, or its employees acting within the scope of their employment, reproduce or distribute one copy or phonorecord of a work under circumstances would typically not amount to fair use. The conditions of the library exemption are that:

- the reproduction or distribution should be made without any purpose of direct or indirect commercial advantage;
- The collections of the library should be open to the public or available not only to researchers affiliated with the library, but also to other persons doing research in specialised field;
- The reproduction or distribution of the work should include a notice of copyright; and
- Specific exemption in subsections (b) through (g) of Section (108) applies (US Copyright Act, 1976,p.84).

The circumstances under which a library may reproduce or distribute a copyrighted work without infringement liability include archival copies. A library may reproduce and distribute a copy or phonorecord of an unpublished work reproduced in facsimile form if the sole purpose is preservation and integrity of the holdings, and if such reproduced copy is currently in the collection of the library.
This exemption does not allow for preservation in electronic or digital form (loc. cit.).

African national libraries and archives are exempted as of Article (14) on Reprographic Reproduction by Libraries and Archive Services, Limitations on Economic Rights in the Agreement Revising the Bangui Agreement of March 2, 1977, on the Creation of an African Intellectual Property Organisation (OAPI), as follows:

- 'Notwithstanding the rules of Article (9), a library or archive service whose activities are not directly or indirectly providing commercial services may, without the consent of the author or other holder of copyright, make individual copies of a work by means of reprographic reproduction;

- Where the work reproduced is an article or a short extract from a written work, other than a computer program, with or without illustration, published in a collection of works or in an issue of a newspaper or periodical, and where the purpose of reproduction is to meet the request of a natural person; and

- Where the making of such copy is for the purposes of preservation and, if necessary, in the event of having lost item, destroyed or made unusable, replacing it or, for replacing a copy that has been lost, destroyed or rendered unusable in the permanent collection of another library or other archive service' (Bangui Agreement, 1977, pp.123-124).
It is noticed that the above Agreement needs to be amended in order to incorporate recent technological advancements such as online library services and preservation of Web-based content programmes for the future African generations.

The philosophy of limitations and exceptions to the exclusive rights granted to authors are looked at through three fundamental freedoms namely freedom of expression, freedom of the press and the right to information (Lepage, 2003). Core of this study is to maintain the right to information that concerns services provided by libraries and archives.

Many digital materials are encrypted as a safeguard against illegal copying and piracy. The American law and the European legislation make it illegal to breach this protection, even if there are good reasons for doing so. Unfortunately the technology makes it possible for a right holder to encrypt content so that users cannot get access to it for fair dealing or fair use (Loffman, 2003, op. cit.).

International copyright rules allow countries to place limits on exclusivity and the right to prevent unauthorised use by permitting reproduction for personal use, research, education and other non-commercial purposes. These exceptions and limitations on exclusivity are helpful to diffusion of knowledge and technology (Michalopouls, 2005, op. cit.).

In relation to the importance of copyright limitations and exceptions, Lung (2004) argues that copyright, like other kinds of IPRs, is basically limited in time, scope as well as exercise. The primary rationale behind such limitations is the need to protect the public interest for citizens to be supplied with information and knowledge, thereby encouraging education, learning and progress of scientific
discoveries. This need should be carefully balanced with the desire to reward authors for their works, and to stimulate them to continue producing new innovative material.

Put simply, King and Mann (2004) believe that there is a way forward to clear the copyright barriers that prevent a blind person in Uganda, for instance, studying English literature, accessing the substantial number of literary publications available in the US or UK in special formats such as audio, custom print or Braille. They believe that the copyright barrier to such service can be resolved in four to five years by coordinating enactment of legislation in each country in accordance with a draft proposed by WIPO for the developing countries.

As discussed above, the copyright exceptions for the visually impaired library users, like other exceptions for the benefit of those with other disability types, constitutes a prominent example of the categories of other library users. Being enshrined in international legal instruments, in particular the UN Universal Declaration on Human Rights (UNUDHR) and the UN Standard Rules on the Equalisation of Opportunity for Disabled People (UNSREODP), this exception aims at securing the right of blind or partially sighted people to access information and knowledge (Lung, 2004, op. cit.).

The essence of exceptions and limitations to copyright is the core of the ‘ratio legis’ or the reason behind the legislation. If copyright protection embodies a monopoly that is entitled to authors for their creative work, such exceptions to these exclusive rights seem to

Latin legal term known as ‘the will of the legislator’ (Peczenik, 1989).
constitute a form of ‘quid pro quo’, allowing users under certain conditions to get access to a work without authorisation required from the owner of the copyright, which shows that in granting the author a monopoly account has been taken of the need to balance the interests of both parties, namely the author and user, which undertakes to protect the author’s creative works (Lepage, 2003, op. cit.p.3).

Dominance and dependency of the proprietary software, as stated in the problems of the study in Chapter (1), section (1.1.2) are created as a result of the monopoly barriers. For this reason, the European Union (EU) has indicted Microsoft in a case of monopoly in the licencing practices the company imposes on consumers. The EU expanded the investigation to specifically examine streaming media technology, or in particular, the Windows Media Player in the Windows platform.

On September 17, 2007, The EU Court of First Instance (EUCFI) upheld the decision reached in 2004 and reaffirmed victory by the EU. The Court found that ‘Microsoft abused its dominant position in the software world to hold back competition in everything from operating systems to server software to media players’, accordingly, it is penalised with the equivalent of a USD$690 million fine, the largest on record ever imposed by the EU (EU v. Microsfot, 2007).

As described previously, any exception for the visually impaired library users should pass the three-step test as outlined in the Berne Convention, the TRIPS Agreement and the WIPO Internet Treaties. The first step appears to be covered, as it applies to certain special cases limited to specified groups of users and

---

51 Latin means ‘something given or received for something else’. Merriam Webster Dictionary (online).
covering certain categories of works and uses. The second step should be satisfied in the sense that such uses may not have the potential to conflict with a normal exploitation of the work. As to the third step, the question of unreasonable prejudice needs to be considered in order to determine if the exception should be subject to a requirement to pay equitable remuneration, or qualified as a free use (Lung, 2004, op. cit.).

IFLA and eIFL (Electronic Information for Libraries), however, made interventions at each of the meetings on the developmental agenda, explaining the role of libraries in the information society to governments representatives, the role of copyright in the provision of library services, how over-restrictive copyright laws can erode access to knowledge and can impede development and why the current ‘one size fits’ approach is unjust and inequitable. eIFL, however, summarises main barriers to library services including how:

- difficult it is for traditional exceptions and limitations to be extended to the digital environment;
- pressure towards increasing terms of protection reduce the public domain upon which opportunities for learning and creativity depend;
- new layers of rights on digital information create barriers and hinder access;
- TPM prevent libraries from availing of lawful exceptions and from preserving national digital cultural heritage;
- unfair and non-negotiable contract terms in licences, which predominantly govern access to digital content, restrict
library statutory rights (eIFL-IP, 2006, pp. 1 of 3, 2 of 3).

Needless to say in this study, the public domain is a primary source of materials that can be preserved without having to face an intricate, potentially expensive, and sometimes hopeless permission clearance process as will be illustrated in the next part.

2.4.4 The Public Domain Information

The ‘public domain information’ is defined as:

'publicly accessible information, the use of which does not infringe any legal right, or any obligation of confidentiality. It thus refers on the one hand to the realm of all works or objects of related rights, which can be exploited by everybody without any authorization, for instance because protection is not granted under national or international law, or because of the expiration of the term of protection. It refers on the other hand to public data and official information produced and voluntarily made available by governments or international organizations’ (UNESCO, 2003, pp.6-7).

The flow of public domain information, argues Guibault (2006), is an essential precondition for cultural, social and economic development. But this right is under threat due to the ongoing race towards the knowledge economy. Information resources, which in the ‘conventional economy’ had little or no economic value have acquired independent economic value in the digital environment, and consequently become subject to IPRs making the information a tradable commodity.
As Boyle (2004) compares, the exponential expansion of IPRs has led to a substantial increase in the imbalance between the public domain and the realm of property rights. Therefore, the basic balance between the public domain and the extent of property rights seems to have been lost though the Web is built on an open source application.

Fundamentally, the Berne Convention has been developed according to the standards and requirements of the industrialised countries in Europe. It has been adopted in September 9th 1886 and revised in Paris in 1971 and administered by WIPO. The minimum standards of protection provided for in the Convention also relate to the duration of protection then works fall into public domain. Article (7) of the Convention establishes a minimum term of protection, which is the life of the author plus fifty years after his death (The Berne Convention, 1886).

Most of countries in the world have legislated for life plus a fifty-year term of protection since it is felt fair and right that the lifetime of the author and the lifetime of his family should be covered, this could also provide the incentive necessary to stimulate creativity, and constitute a fair balance between the interests of the authors and the needs of society. The term of protection, in so far as moral rights are concerned, extends at least until the expiry of the economic rights (Ibid.).

Nevertheless, as part of public domain, e-Government publications play a fundamental role in advancing activities of the e-governing. National libraries play a major role as memory organisation equipped with legal instrument of legal deposit to collect and manage them (Cruse, 2003).

During the last ten years Nicholson (2007, op. cit.) enumerates the impacts that face library
services arguing that the international trend in
copyright denied the right to the public domain
information, restricted access to knowledge and
tightened protection of digital content in
particular. International IPRs treaties, EU
Directives and national legislations have eroded
the right to information.

Erosion of access to information for future
access basically contradicts provisions stated in
Article (19) of the UNUDHR which states that:

‘Everyone has the right to freedom of
opinion and expression; this right
includes freedom to hold opinions
without interference and to seek,
receive and impart information and ideas
through any media regardless of
frontiers’ (UNUDHR, 1948).

According to the IFLA Internet Manifesto,
unhindered access to information is essential to
freedom, equality, global understanding and
peace. Therefore, the Manifesto (2002, op. cit.)
asserts that:

- Intellectual freedom is the right of
every individual both to hold and express
opinions and to seek and receive
information, it is the basis of
democracy, and it is at the core of
library service.

- Freedom of access to information,
regardless of medium and frontiers, is a
central responsibility of the library and
information profession.

- The provision of unhindered access to the
Internet by libraries and information
services supports communities and
individuals to attain freedom, prosperity and development.

• Barriers to the flow of information should be removed, especially those that promote inequality, poverty, and despair.

Non-governmental organisations such as IFLA and eIFL also advocate the free access to information and knowledge as a priority for participation in the information society without geographic, cultural, social and regional obstacles.

As of Besek’s (2003) arguments, there are three international issues should be considered in planning a digital archive. First, international treaties place certain constraints on countries’ ability to specify exceptions to copyright protection or to impose requirements on copyright holders. Second, legal and logistical uncertainties could make it difficult for a copyright owner to obtain redress for copyright infringements committed abroad. These uncertainties should be considered in deciding which works should be included in the digital archive and from where they will be accessible. Third, a digital archive that permits online access outside the country could itself be vulnerable to suit by foreign copyright owners whose works are included. Also foreign attacks of intruders, as discussed later, remain major threat.
2.4.5 Threats of Hacking and Piracy

For the purposes of this study, it is important to talk about the case of IPRs infringement and piracy, i.e., ‘prima facie’ [based on a first impression] evidence. Hackers and pirates, for instance, crack the code of e-books or other material and then put the item on a peer-to-peer online group then the network grows exponentially as friends tell friends on the cyberspace. Consequently, laws are tightened and exemptions evaporate. According to Idris (2007), the WIPO Director, trans-border counterfeiting and piracy cause economic loses estimated at more than USD$150 billion annually.

For copyrighted materials laws reserve certain separate and exclusive rights to copyright holders and their licencees. These include, according to Hanratty (2005), the right to make and distribute copies of a work, the right to prepare derivative works based upon the copyrighted work, and the right to display and perform the work to public. But to prove infringement a copyright holder only needs to show the following:

- ownership of the copyright in the work; and
- proves that original elements of the work are copied.

It is not necessary to prove that the copy is in the same medium as the original, as long as it is linked and communicated to others. If such requirements are fulfilled, then the breach has taken place and liability can only be negated if the violator can offer a valid defence. But a challenge remains viable in the case of linkable content on the Web. Moreover, the most intricate legal problem is how to trace liability when a central server cannot be identified due to
software being untraceable where there is no need for central server; the sheer volume of participants makes any kind of legislation impotent (Hoorebeek, 2003).

The issue of piracy raises questions on the legality of Google Library project (see also section 2.3.7). In October 19, 2005 five publishers including McGraw-Hill, Pearson, Penguin, Simon and Schuster, and John Wiley and Sons sued Google, Inc. Similarly, in September 20, 2005 the Authors Guild and several individual authors sued Google, Inc. for copyright infringement. It is concluded that the authors claimed damages and compensation. While publishers, in contrast, only requested compensation. Google’s philosophy is based on the assumptions that the copying it maintains is legal under the fair use doctrine\(^{52}\) pursuant to Article (17) Section (107) of the US Copyright Act of 1976 (as amended in 2007; see also section (2.6.1) with the aim to ease and locate information buried in books and the custody of libraries (Band, 2006, op. cit., pp.3-4).

In its defence, Google Library Project depended on the premises that usage of a copyrighted material that would otherwise breach the copyright owner’s exclusive rights is not a contravention if it is determined to be fair use. Section (107) of the Act, as referred to above, states that four non-exclusive factors to be considered in fair use analysis on a case-by-case basis as follows:

\(^{52}\) It is a privilege originally created by judges in the 19\(^{th}\) century. It has been subsequently made a part of the US Copyright Act when approved in 1976. Section (107) provides that “the fair use of a copyrighted work... for purposes such as criticism, comment, news reporting, teaching, scholarship, or research is not an infringement of copyright” (Fishman, 2008, p.292).
• the purpose and character of use (including if it is commercial in nature or a ‘transformative’ use;
• the nature of the copyrighted material;
• the amount of the material used; and
• the effects or potential effects on the market for the original work (US Copyright Act, 1976).

In 17 February 2006, the Google Project has been found guilty in Perfect 10 (Plaintiff’s) v. Google, Inc. case. The District Court for Central District of California established that Google’s creation and public display of thumbnails likely did infringe the Plaintiff’s copyright protection and did not constitute fair use (District Court Perfect 10 v. Google, Inc., 2006).

Also Viacom International and its companies have filed a copyright infringement lawsuit against YouTube and Google seeking at least USD$1 billion in damages. The media company charges that ‘YouTube has harnessed technology to wilfully infringe copyrights on a huge scale’, through taking ‘the value of creative content on a massive scale for YouTube’s benefit without payment or licence. The suit alleges that the copyright infringement is on such a large scale that it ‘identified more than 150,000 unauthorised clips of their copyrighted programming on YouTube that had been viewed an astounding 1.5 billion times’ (Viacom et al. v. Youtube et al. 2008).

The Internet Archive is also criticised and sued by Healthcare Advocates, which comments that ‘It is just like a big vacuum cleaner, sucking up information and making it available’. If it is not settled out of court, this will be an interesting case for more digitally adventurous preservation projects to learn from. This is because the Internet Archive did not ask
copyright clearance before preserving the entire Internet, although it responds to requests to restrict content (Bailey, 2006, op. cit., p.124).

Similarly, in December 1999, the Recording Industry Association of America (RIAA) launched legal proceedings against Napster, Inc., a company founded by Shawn Fanning in the same year, and worked to facilitate the sharing of computer files through the Internet. The company has been indicted for enabling clones or file sharing of DVD\textsuperscript{53} music on the Web and accordingly it has shut its doors in May 2002 after intense legal disputes in the US courts (Hoorebeek, 2003, op. cit.). A survey made in August 2001 by Envisional, a British right management company, claimed that 7,300 print editions of popular books have been scanned and made available on the Web through Napster clones (Godwin, 2001).

Due to the clones made by Napster, as above-mentioned, the publishing industry shares their concern over the Web publishing. On August 28, 2001 a Federal Grand Jury has indicted the Russian company Elcomsoft and programmer Dmitry Sklyarov on charges of trafficking and conspiracy to traffic in a copyright circumvention devices. These charges were presented for infringement of the Digital Millennium Copyright Act (DMCA) of 1998. Elcomsoft has breached the DMCA pursuant to the US Law when sold software that disables the encryption of Adobe e-books. But the Jury neither prosecuted the company nor the programmer for manufacturing software that hackers and pirates might utilise to crack online databases (Glasner, 2002).

The researcher argues that Elcomsoft case indicates that Web-based content is uncontrollable through enforcement of national IPRs over trans-border jurisdiction. This is because there are no international binding IPRs

\footnote{53 Digital versatile disc.}
to govern the relationship of flow of intellectual products between nations. Furthermore, encryption methods are complex and difficult task. As soon as one programmer creates a new software encryption device, the programmer’s wits are matched against the countless hackers who can crack the code and post a copy in a peer-to-peer (P2P) file-sharing mode. Another area of concern on the Web environment where most content is automatically copyrighted, copyright terms are lengthy or endless, and content is increasingly licenced rather than owned in contrast to print library items. Simply put, a national library cannot preserve what it does not own unless the work fall into the public domain, the work’s licence permits it, the work’s copyright owner grants permission to do so, or Web legal deposit enforces such act as proposed in Chapter (4) section (4.5). There are other challenges such as the tendency towards commercialisation of information also contribute to the problems of preservation as seen in the following section.

2.4.6 Implications of the GATS and TRIPs

The international community developed treaties that ensure protection for intellectual property such as the Trade-Related Aspects of Intellectual Property (TRIPS) introduced in 1994 (Matsika, 2007). The main effect of TRIPS Agreement is that it imports most of the provisions of the Berne Convention for the Protection of Literary and Artistic Works (IFLA-CCLM, 2002).

Another legal hurdle related to intellectual property is the General Agreement on Trade in Services (GATS). The World Trade Organisation (WTO) administers both TRIPS and GATS agreements. On 29 November 2005 the Council for TRIPS extended, until 1 July 2013, the transition period for the Least-developed countries
(including the Sudan) to implement the TRIPS Agreement. A key element of the decision is related to technical and financial cooperation (Musungu, 2007).

Upon the significance of the WTO decisions on libraries, IFLA (2001) is concerned that a growing evidence that the WTO decisions may affect the prospective development of library operations and services.

Though TRIPS agreement aims to develop innovation through fair competition and protection of rights, some of the least developing countries’ concerns about the implications of stricter implementation of the IPRs for their economies, and in particular, for transfer of technology, have received limited attention (Correa, 2007). The duration of copyright protection of works, according to the TRIPS Agreement, is the life of the author plus fifty years (IFLA-CCLM, 2002, op. cit.).

Reference to Rikowski (2003), implementation of the TRIPS agreement will have a significantly more detrimental impact on the least developed countries than on the developed one, at least in its initial implementation phase. Nicholson (2002, p.261) supports Rikowski’s point of view saying that due to ignorance of their intellectual property and other rights, rural people are often at the mercy of large transnational companies and individuals who recognise the profits in their traditional remedies, music, folklore, craftwork and other conventional cultural works. These people, Nicholson adds, are not aware of the legal requirements of having to put their oral expressions or traditional intellectual property into a tangible format, before they can claim copyright ownership. But without access to information, they are unaware that their rights
are misappropriated and used for commercial exploitation in other countries.

Another indirect impact of TRIPS, as Rikowski (2003, op. cit.) points out, is that the high rates of illiteracy in many African countries constitute a barricade that intensifies the problem of misappropriation. Much local knowledge is passed down from one generation to the next verbally and is not recorded in any tangible format. She also argues that the GATS agreement is targeting fairly on liberalising trade in services in the marketplace and accordingly this threatens the state-funded provision of libraries.

Since African communities are basically dependant on agricultural production, there are adverse impacts of the TRIPS on the technology transfer process as summarised as follows:

- Privatisation of information leads to less reliance on extension and more dependence on private technical consultants.
- Lack of public sector access to information internalised by individuals and firms will curtail the ability to conduct resource inventories, detect market failures, and undertake policy analysis.
- Fewer farms with a concentration of assets will enjoy differential access to world-class-information, leading to a geographical concentration of production in areas where the information infrastructure is well developed (Reddy, 2007).

Also there is some concern about healthcare projects aiming to combat diseases such as
malaria, tuberculosis, and HIV/AIDS\textsuperscript{54} in Africa as due to lack of access to information for health awareness though some initiatives plan to achieve the goal of ‘Information For All by 2015’ (Pakenham-Walsh, 2007).

Nonetheless, there is uncertainty about what ‘services’ will fall under the GATS. The term is defined in the GATS Agreement in Part (1) Article 1(b) as follows:

\begin{quote}
‘services’ includes any service in any sector except services supplied in the exercise of governmental authority’ (GATS, 1994,p.286).
\end{quote}

In the above definition it is likely that national libraries and archives are excluded. Furthermore, Part (1) Article (C) states that:

\begin{quote}
‘a ‘service’ supplied in the exercise of governmental authority means any service which is supplied neither on a commercial basis, nor in competition with one or more service suppliers’ (Ibid.).
\end{quote}

However, there are few public services today that have no element of competition at all. Thus, libraries are left with a little room for interpretation. Whitney (2000,p.1), who represented IFLA at the WTO Ministerial Conference in Seattle in 1999, notes that key negotiators have made it clear in Seattle that all services are subject to GATS revision, including education, healthcare and indeed, libraries, archives and museums and that hard

\textsuperscript{54} Acquired Immune Deficiency Syndrome/Human Immunodeficiency Virus.
choices should have to be made which may include the sacrifice of some sectors to protect others.

The major impact of the GATS and TRIPS agreements on preservation of library material is that they slowly ebbing away material in the public domain. Arguably, Rikowski (2005) warns that the agreements are potentially affecting the survival of public libraries in particular and to access to information in general because they aim to proceed with the WTO’s neo-liberal agenda by transforming IPRs, through the TRIPS, and services, through the GATS, into global tradable commodities. In other words, the agreements are widening the imbalance between the right to protect interests of the rightholders against the public interests of information needs. Consequently, libraries are prone to privatisation and competition.

Implementation of these agreements has been widely criticised for over emphasising the benefits of intellectual property for rightholders while paying little attention to the costs, for being partisan and for encouraging a so-called TRIPS-plus approach (eIFL-IP, 2006, op. cit.).

Researches (Musungu & Dutfield, 2003, op. cit.) found that many of the rules and or principles embodied in the TRIPS agreement existed in some way or another in various treaties administered by WIPO. Although the agreement introduced significant changes in the overall framework of the international IPRs regime, it has not in fact altered the standard setting structure. In other words, the agreement has substantially changed such regime by introducing the principle of minimum IPRs standards. In effect, this doctrine means that any IPRs agreement negotiated subsequent to TRIPS among and or involving WTO members can only create higher standards (commonly known as ‘TRIPS-plus’). The arrival of
TRIPS has ushered in an era of peace for WIPO concerning the development-related demands by developing countries. For WIPO itself, the advent of TRIPS created a significant strategic dilemma. This is because WIPO had to share its hitherto ‘exclusive competence’ on IPRs issues with the WTO.

In September 2004, Brazil and Argentina have made a historic proposal to establish a ‘development agenda’ within WIPO. The Agenda aims to re-orient WIPO to its original goal to promote intellectual creativity rather than intellectual property. Joined by Bolivia, Cuba, Dominican Republic, Ecuador, Egypt, Iran, Kenya, Peru, Sierra Leone, South Africa, Tanzania and Venezuela, the group came to be known collectively as ‘Friends of Development.’ Their overall aim is to promote development and access to knowledge for all (eIFL, 2006, op. cit.).

It is noteworthy that IFLA and eIFL have been the early signatories to the Geneva Declaration on the Future of WIPO adopted by leading academics, Nobel Prize winning scientists, access to medicine advocates and free software developers (Geneva Declaration, 2004).

TRIPS-plus includes any new standards that would limit the ability of least developed countries to:

- promote technological innovation and to facilitate the transfer and dissemination of technology;
- take necessary measures to protect public health, nutrition and to promote the public interest in sectors of vital importance to their socio-economic and technological development; or,
- take appropriate measures to prevent the abuse of IPRs by right holders or the resort by right holders to practices
which unreasonably restrain trade or adversely affect the international transfer of technology (Musungu & Dutfield, 2003, *op. cit*).

As a result of the above-mentioned standards, the concept of minimum IPRs standard covers both those activities aimed at increasing the level of protection for right holders beyond that which is given in the TRIPS agreement and those measures aimed at reducing the scope or effectiveness of limitations on rights and exceptions under the TRIPS agreement (*Ibid*.).

Evaporation of such rights directly affects library ability to preserve digital content in general and the Web-based one in particular. Reducing the exceptions by the TRIPS agreement also results in cutting library funding. In this sense, the worst scenario is that libraries could disappear or, in Hunt’s (2001) terms, other competing information companies might replace libraries by way of privatisation. Accordingly, the public would then be required to buy their information from information companies or from libraries, if such libraries could remain afloat by providing information services for fees. Either way, the public would find themselves paying for information that once available in the public domain.

In a subsequent statement, however, IFLA criticised WIPO for failing to protect and promote the balance between IPRs owners and library users (IFLA, 2004, *op. cit*.).

Movement on access to knowledge has emerged and has been growing fast as advocated by the ‘Friends of Development’ discussed above. A treaty on access to knowledge for developing countries has been proposed to address at least two concerns:
• those relating to investments in human capital particularly education and health;
• those regarding models that maximise the participation of developing countries in the processes of innovation and the spill over benefits of knowledge while minimising the social cost of accumulating knowledge (Musungu, 2005).

The writer recommends that such proposed treaty should be administered by WIPO only in response to his opposition of the WTO role in this concern. There are other IPRs treaties that will be investigated in the next part.

2.4.7 The WIPO Internet Treaties

The invention of the Web has added the capacity for massive reproduction and distribution of content. This is an issue that previously seen as a matter between giant industries but currently reversed into an issue that affect individuals in their everyday life. The potential for the Web and its ability to facilitate growth and development is just beginning to be understood. In the IPRs debate, the amazing ability of this tool has largely been featured as increasing the threat of piracy of such property (Musungu, 2005, op. cit. p.9).

The WIPO Internet Treaties are key instruments on the WIPO Digital Agenda approved by the Member States in September 1999. It aims to safeguard global protection for creativity (WIPO, 1999, op. cit.).

The Treaties introduced new TPM rights management measures through a set of obligations for the countries that adhere to. These obligations are designed to ensure that rightholders may effectively use technology to protect their rights and to licence their works
online. The first obligation requires countries to provide adequate legal protection and effective remedies against the circumvention of technological measures, such as conditional access systems and encryption used by right holders to protect their property. The second type of technological safeguards enhance the reliability and integrity of the online marketplace by requiring countries to prohibit the deliberate alteration or deletion of electronic information which accompanies any protected material, and which identifies the work, right owners, and the terms and conditions for its use, among other things (Lung, 2004, op. cit.).

The WIPO Treaties gave expression to the desire of its signatories to tighten the copyright and related rights restrictions while reaffirming the need to ensure a balance between the latter and the general public. The intention has also been a response to the critics who contested the monopolies emerging from national IPRs laws from the point of view of development of a global cyberspace. It is true that the Internet initially seemed to be ‘a legal no-man’s-land’, open and uncontrollable, whether by the government authorities or by the information technology professionals, until national courts established a package of Internet laws come out of legislation applicable to analogue media (Gaubiac, 2002, op. cit.).

The most important fact reached is that access to information under the WIPO Internet Treaties is threatened. This is because what is possible for a user in an analogue environment is not possible on the Web platform. The concept of fair use or fair dealing limitations on the Web might disappear. For example, if technical anti-circumvention measures applied, the quotation of work, pastiche and its utilisation for
educational purposes might be prevented (Lepage, 2003). The coming section compares between two different principles to the rights of preservation of library materials.

2.4.8 Legal Deposit v Copyright Deposit

Preservation of national library and archives materials is often enshrined under national legislation of a certain country for maintaining longevity and flow of cultural heritage from generation to another. The national legislation differs from country to country and ranges from legal deposit, library deposit, and voluntary deposit to copyright deposit.

2.4.8.1 The Legal Deposit

The main purpose of the legal deposit provisions is characterised as to develop and preserve the nation’s published heritage and make them accessible to present and future national library users, in the case of this study, to establish a national Web archive (Lupovici, 2005; Muir, 2004).

Legal deposit also aims at the enrichment of national library material and it is an instrument used to collect a full and permanent record of the nation’s published works and of a record of all disciplines of knowledge limited to publications within national boundaries (Gilchrist, 2006). Researches (Lupovici, 2005, op. cit., p.2) state that legal deposit is more a question of type of content and or distribution channels (books, engravings, maps, multimedia, audiovisuals and others) than a question of media such as paper, disc or online. On the Web, most of institutional grey literature will be part of the legal deposit domain as long as documents on the Web are accessible to the public, thus published.
The national boundary limit assumption is no longer applicable in the case of preservation on the Web. This is made clear by the American Library Association (ALA) when launched strategy for 2010 stating that the strategy will increase influence in promoting the preservation of the American cultural heritage (ALA, 2005).

Legal deposit of digital materials requires copying from online and deep Web, i.e., databases, or surface Web-based content. As a consequence, there is a need to define what the legal deposit is. The UNESCO defines the ‘legal deposit’ as follows:

'A statutory obligation which requires that any organization, commercial or public, and any individual producing any type of documentation in multiple copies, be obliged to deposit one or more copies with a recognised national institution' (UNESCO, 2000,p.3).

This definition asserts that it is important to make sure that legal deposit legislation covers all kinds of published material, that is, material generally produced in multiple copies and provided to the public regardless of the means of transmission, in order to differentiate from ‘archival’ which refers to records, either governmental, corporate or personal and which are usually unique items, not available for public distribution and more of a private or personal nature though some of these boundaries have been eliminated by the Web.

The British Library (2005) defines the law as: 'The act of depositing published material in designated libraries'. This definition seems simple and assigns ‘libraries’ to act as
official depository institutions. While the UNESCO’s definition has put legal deposit responsibility more generally to any ‘national’ and ‘recognised institution’ that may include national archives, national libraries or other institutions.

PADI, also defines it as follows:

‘Legal deposit (also termed ‘dépôt légal’ or ‘mandatory deposit’) is a statutory provision which obliges publishers to deposit copies of their publications in libraries in the country in which they are published. The principle of legal deposit is established in international convention and in the national legislation of many countries, and aims to ensure that access to a nation’s published cultural material in libraries and archives is preserved’. (PADI Website).

As of this study, PADI’s philosophy is that if legal deposit is implemented to include digital information, the protection of publishers’ rights and investments needs to be considered. This is because dissemination of information on the Web poses further challenges because there is no physical equivalent to deposit.

The Council on Library and Information Resources (CLIR) emphasises the importance of legal deposit in preservation of scientific journals commenting that although legal deposit may not be the silver-bullet solution to preservation of e-journals, it is obviously an important component of the preservation equation. If nothing else, a legal requirement that would force publishers to deposit e-journals in several national deposit agencies
would create pressure for standard submission formats and manifests for e-journal content. In addition, once material is preserved, it may be possible to revisit the trigger events that allow access to the content and even to permit remote access in certain situations (CLIR, 2006, pp. 22-23).

Though there is slighter ambiguity regarding the term ‘published’ on the Web, but it is clarified that a radio or television programme, for instance, could be treated as ‘published’ for legal deposit purposes when it has been broadcast such as the Finish law which obliges deposit of digital publications as well as radio and television material (Wiggins, 2005).

Furthermore, in the digital publications environment, it should be noted that a ‘one copy document,’ of a database stored on one server, could be subject to a legal deposit requirement since it is made available to the public through network enabling the public to read, hear or view the material (UNESCO, 2000).

Generally, any library material is an object of legal deposit as long as it is made available to the general public and produced in multiple copies. Some studies enumerate items covered by legal deposit as all types of printed material such as books, serials, pamphlets, maps, manuscripts, and other materials, to most audio-visual items such as discs, films, videos, multimedia kits, and others, to broadcast programmes and to electronic objects such as diskettes, CD-ROMs, online and other material (Husband & Green, 2008).

The principle of a legal deposit was first implemented in 1537 when King Francois I of France issued the ‘Ordonnance de Montpellier’\(^{55}\). It was a royal decree, which

\(^{55}\) Montpellier Ordinance of 28 December 1537.
prohibited the sale of any book without first having deposited a copy in the library of his castle (Lupovici, 2005, op. cit., p.2).

The King’s objectives were to collect and gather the existing and future production of all editions of the books that ‘deserve to be seen’ to enable him to ensure that it would often be possible to refer to the original work as ‘first published and not modified.’ Though it was officially entitled royal status, historians had reported that the decree was not well respected, but it is essential to emphasise that the principle had been established and implemented in the majority of other countries of the world (UNESCO, 2000, op. cit.).

2.4.8.2 Copyright Deposit

Provisions of the legal deposit was first abolished during the French Revolution, in the name of freedom, and were reinstated in 1793 as a formality to obtain copyright protection (UNESCO, 2000, op.cit.). In this sense such formality is considered hereinafter as ‘copyright deposit.’

As early as 1594, Belgium had implemented a legal deposit system, but it was abolished when the Berne Convention, the first international treaty on copyright, had been enacted in 1886. Most other countries adopted legal deposit, but not as a copyright formality or deposit, whereas Belgium just abolished it and reinstated it in 1966 (Jasion, 1991). But the principle of ‘no formalities’ and the minimum duration of fifty years after the author’s death were some of the reservations for the US to adhere to the Berne Convention and this, in turn, led to establishment of the unbinding Universal Copyright Convention (UCC) in 1952 which revised in 1971 (Lewinski, 2006).
During the seventeenth century, the principle of legal deposit was expanded when Ferdinand the Second, a Germanic emperor from 1619 to 1637, ordered in 1624 that one copy of each book published be sent to the library of his court (Gilchrist, 2005, op. cit.).

The relationship between legal deposit and copyright has strengthened during the eighteenth century when legal deposit became the formality for obtaining legal protection of copyright. This began when Statute of Anne, the Great Britain Copyright Act of 1709 and the first legislation that aimed at protecting authors’ works from piracy, was established. It required that nine copies of work be deposited and distributed to several libraries in order to obtain copyright protection (Albert et al. 1999, op. cit. p.208).

Most countries had modified their legal deposit legislation when the Berne Convention was implemented in 1886. The Convention clearly stated that the enjoyment and the exercise of the right to protection of literary and artistic works should not be subject to any other legal formality, consequently, deposit as a requirement to obtain copyright protection was abolished. Except for a few countries, such as Belgium, most countries retained deposit as a legal requirement through another means, such as a specific law on legal deposit. The Institut international de cooperation intellectuelle de Paris (IICIP) conducted study in 1938, which identified 52 countries that had an enforced legal deposit law, through a law, or another legal means such as a decree, an edict, an order or a regulation. Fifty years later, in 1990, another study surveyed 139 countries having a formal legal deposit law in one form or another (Loc cit. p.8).
One important relationship between legal deposit and IPRs is that all works protected by the later in general, and the copyright in particular are subject to legal deposit in national libraries. It is apparent that the legal deposit is significant in building collections and national bibliography in national libraries (Wiggins, 2005, op. cit.).

Legal deposit is a vital instrument for preservation programmes but is quite separate from copyright legislation in most countries. The ability to copy something for preservation is determined by certain criteria such as the age of the material, its format and the reasons for copying. The different processes of photocopying, microfilming and digitisation, all pose specific legal questions and also create difficulties. The end product of preservation project themselves can be subject to copyright. As the use of digital media grows so the legal issues surrounding preservation need to be reviewed carefully (Cornish, 1994).

In the researcher’s opinion, looking to the copyright deposit in another perspective, after the enforcement of the Berne Convention, the divorce had taken place between the copyright deposit and the legal deposit. As already mentioned, in many countries the early deposit of national library collections were built up through the copyright deposit, whereby copyright protection could not be claimed unless copies of the item had been deposited as the case in the Sudan National Library. Hence national libraries rely on its provisions for reinforcing the process of building their holdings.

For the sake of comparison, both the legal deposit and the copyright deposit aim at two different ends. The later aims at protecting economic and moral rights of authors and
creators, while the former aims at protecting the right of access to information through preservation. But this did not prevent some countries from including legal deposit provisions within its copyright legislation, as in the US, the UK and other countries.

It is evident that IPRs constitute major barriers to library services generally and to the preservation projects in particular. In the next part, the study links between the IPRs and the role of the international organisations designated responsibilities to administer them.

2.5 Role of the International Organisations

The international community represented in the UN, has established international organisations in order to reinforce rule of law regarding protection of IPRs. Theses organisations are:

2.5.1 WIPO

The establishment of secretariats in the form of international bureaux accompanied the adoption of both Paris and Berne Conventions. The two bureaux were merged in 1893 to create the Bureaux Internationaux reunis pour la protection de la propriete intellectuelle (BIRPI), the immediate predecessor of the World Intellectual Property Organisation (WIPO). BIRPI, which was originally based in Berne before moving to Geneva in 1960, was responsible for administering both the Paris and Berne Conventions in addition to a number of special agreements under the Paris Convention (Musungu & Dutfield, 2003, op. cit).

The WIPO, therefore, is established in Geneva in 1970 when the Stockholm Convention came into force, and subsequently became a specialised agency of the UN in 1974 with a mandate subject
to a number of other UN organisations dealing with innovation, development and intellectual property (Ibid.).

The strategic goal of the Organisation for 2006-07 is to ‘promote an intellectual property culture’. It does not dependant, in its funding, on contributions of member states as other UN agencies but instead gets 90% of its income from the collection of fees under the patent registration scheme which it administers. In other words, right holders, who display a keen interest in expanding intellectual property protection, largely fund the Organisation. In a cooperation agreement with the WTO, the WIPO Secretariat provides technical assistance and legislative advice to developing countries on national implementation of the TRIPS agreement (eIFL, 2006, op. cit.).

WIPO administers various international agreements such as Berne Convention for the Protection of Literary and Artistic Works and the Rome Convention for the Protection of Performers, Producers of Phonograms and Broadcasting Organisations, and more recently WIPO Copyright Treaty (WCT) and the WIPO Performances and Phonograms Treaty (WPPT), also known as the WIPO Internet Treaties (Lung, 2004, op. cit.).

By becoming members of WIPO, 171 States have subscribed to the importance of promoting the protection of intellectual property. Many of these have also adhered to some or all of the 16 other multilateral treaties administered by WIPO, which establish international frameworks for each of the rights that make up intellectual property or systems for obtaining protection in multiple countries (WIPO, 1999).
2.5.2 WTO

The World Trade Organisation (WTO) is founded in Geneva in 1995. One of the youngest of the international organisations, the WTO is the successor to the General Agreement on Tariffs and Trade (GATT) established in the wake of the Second World War. Though the WTO is still in its infancy, the multilateral trading system that has been originally established under GATT is well over 50 years old. It is one of two international organisations, which are involved in setting the international copyright agenda (eIFL-IP, 2006, op. cit.).

The system of the organisation has been developed through a series of trade negotiations, or rounds, held under GATT. The first rounds dealt mainly with tariff reductions but later negotiations included other areas such as anti-dumping and non-tariff measures. The last round which held in 1986-94 in the Uruguay Round, led to establishment of the WTO (Reddy, 2007, op. cit., p.2).

The negotiations have not ended there. Some continued after the end of the Round. In February 1997 agreement has been reached on telecommunications services, with 69 governments agreeing to wide-ranging liberalisation measures that went beyond those agreed in the Round (Ibid.).

In the same year 40 governments successfully concluded negotiations for tariff-free trade in information technology products, and 70 members concluded a financial services deal covering more than 95% of trade in banking, insurance, securities and financial information (Ibid.).

In addition, the 134 States that are members of the WTO have subscribed to a comprehensive, complementary code of IPRs protection in the
In 2000, new talks started on agriculture and services. These have currently been incorporated into a broader agenda launched at the fourth WTO Ministerial Conference in Doha, Qatar, in November 2001 (WTO Website).

The work programme, the Doha Development Agenda (DDA), adds negotiations and other work on non-agricultural tariffs, trade and environment, WTO rules such as anti-dumping and subsidies, investment, competition policy, trade facilitation, transparency in government procurement, intellectual property, and a range of issues raised by developing countries as difficulties they face in implementing the present WTO agreements (Ibid.).

The Organisation administers and enforces the multilateral trade agreement that brought copyright into the global trading system. By December 2005, 149 countries had joined the WTO and subsequently TRIPS as discussed in section (2.4.6). The duration is extended for the Least Developed Countries until July 2013 to accede to TRIPS (eIFL-IP, 2006, op. cit.).

An important part of the WTO treaty system is the Agreement on Trade-Related Aspects of IPRs known as TRIPS. This agreement is intended to set common standards for the IPRs regimes of all member states. It covers the following categories:

- Copyright and related rights
- Trademarks, including service marks
- Geographical indications
- Industrial designs
- Patents
- Layout-designs (topographies) of integrated circuits
• Undisclosed information, including trade secrets (IFLA-CCLM, 2002, op. cit.).

In the classification of acceding countries of the WTO, there are only three Arab countries namely Saudi Arabia in 2005 and Oman in 2000 as developing countries and Jordan in 2000 as the least developing one. In Africa, only Angola (1996) has acceded to the organisation so far (Abbott & Correa, 2007, op. cit., 38).

The WTO resolutions are influenced by the so-called QUAD nations, the US, Canada, Japan and the European Union. The Organisation strives to abolish the public sector and encourage privatisation and deregulation. Therefore, it has taken away the right to choose in favour of the right to trade (Hunt, 2001, op. cit.).

2.5.3 ARIPO

The history of the African Regional Intellectual Property Organisation (ARIPO) goes back to the seventies when a Regional Seminar on Patents and Copyright for the English-speaking African countries was held in Nairobi. ARIPO was established after the Lusaka Agreement on the Creation of the Industrial Property Organisation for English-speaking Africa (ESARIPO) on December 9, 1976. The organisation has been mainly established to pool the resources of its member countries in industrial property matters together in order to avoid duplication of financial and human resources (ARIPO Website).

Thus the preamble to the Lusaka Agreement clearly states that member states are:

'aware of the advantage to be derived by them from the effective and continuous exchange of information and
harmonization and co-ordination of their laws and activities in industrial property matters’ (Lusaka Agreement, 1976).

Member states also recognised that the creation of the organisation for the study and promotion of and co-operation in industrial property matters would best serve that purpose (ARIPO Website, op. cit.).

Though the Lusaka Agreement has established the ARIPO, but it has not clarified its powers and functions as an industrial property regional organisation. Therefore, two additional legal instruments have been adopted mainly the Harare Protocol on Patents and Industrial Designs in December 1982 and the Banjul Protocol on Trademarks which have come into force on March 6, 1997 (ARIPO/WIPO Website).

2.5.4 OAPI

The African Intellectual Property Organisation (OAPI) has been established according to Libreville Agreement on September 13, 1962 (OAPI Website). Article (22)2 of the organisation states that:

‘Any African State that is not party to the Bangui Agreement but is party to the Convention Establishing the World Intellectual Property Organization, the Paris Convention for the Protection of Industrial Property, the Berne Convention for the Protection of Literary and Artistic Works and or the Universal Copyright Convention, and the Patent Cooperation Treaty may accede to this Agreement’ (Bangui Agreement, 1977, op. cit.).
This Article illustrates that the international legal system on IPRs are based on a hierarchical structure. As a result, establishing or amending national legislation needs to consider such framework to assist in building a harmonised legal balance. The next part is a voyage to review legal deposit update in major national libraries of the world.

2.6 Global Update of Legal Deposit

During the last ten years, national libraries and other custodial institutions have had some difficult decisions to make regarding the extension of legislation to cover deposit and exemption of digital material as well as Web-based content. The coming part investigates the role of the most national libraries in this arena.

2.6.1 The Library of Congress

Though parliamentarian in its nature, the Library of Congress plays the role of the national library of the US. In the same time, there are two other special libraries acting as national libraries such as the National Library of Medicine and the National Library of Agriculture. The first US Copyright Act was enacted in 1790, it protected published and unpublished library materials written by living national authors (Bailey, 2006, op. cit., p. 117). Through the Copyright Act of 1870, the television, film, and sound recording preservation acts, and most recently, the Veterans History Project, the US Congress has anticipated and addressed the challenge of preserving America’s cultural heritage in its many forms (NDIIPP, 2002, op. cit.). For the meantime, copyright protects a wide range of
published and unpublished works (Bailey, 2006, op. cit., p. 117).

In the US, there is stricter tool for protecting government content in the digital environment known as DMCA of 1998 as mentioned previously in section (2.4.5). It criminalises any infringing conduct (Campbell, 2005). Though this trend, there is a law of Freedom of Information Act (FOIA) of 1966 which enacted by President Lyndon B. Johnson the 36th President of the US Thirty years later, President William J. Clinton, the 42nd president of the US, enacted the Electronic Freedom of Information Act of 1996 bringing FOIA into the digital age. The law enables the citizens to freely access government information as a cornerstone of democracy (Oltmann et al. 2006).

A requirement for legal deposit in the US was first established as part of the 1790 Copyright Act provisions (UNESCO, 2000, op. cit. p.7).

Exceptions for visually impaired library users are enshrined in Section (121) (the Chafee Amendment of 1996) of the US Copyright Law (Lung, 2004, op. cit.). As discussed previously in section (2.4.3.2) in this Chapter, the US Copyright Act of 1976 and its amendments do not allow exemptions for preservation of electronic or digital form. Notwithstanding, the Internet Archive and MINERVA have established and managed Web archives and collected foreign materials served as record for war on terror (see Chapter (1), section (1.9.1.1). The Library of Congress', which was established in 1800 by John Adams, policy towards dealing with the IPRs challenges recommends the following actions:

- Investigation of the implications of mandatory deposit for digital content.
• Investigation of the implication of various security and protection devices for preservation, and
• Development of a better understanding of the international context of copyright, jurisdiction, responsibility, and reach of applicable law, possibly in co-operation with other national libraries and multinational publishing and media industries (NDIIPP, 2002, op. cit., p.6).

It should be noted that the US Copyright Office has been part of the Library of Congress since 1870. In addition to administering the copyright law the Office creates and maintains the public record of copyright registration, provides technical assistance and policy advice on copyright issues to the Congress and executive branch agencies, enables access to information for the general public, promotes improved copyright protection for the creative works conducted abroad, and obtains deposits for the collections of the Library of Congress (The Library of Congress, 1996).

2.6.2 The Library and Archives of Canada

The Canadian Government Web Archive has been established after amending section (82) of the 2004 Library and Archives Canada Act. This amendment broadened scope of the legal deposit to allow capture and harvest of Web-based content since early 2007 (Cantello & Stegena, 2008, pp.1, 12).
2.6.3 The British Library

The legal deposit law in Britain was introduced in 1610 when Sir Thomas Bodley made an agreement with the Stationer’s Company. Pursuant to the terms of that agreement, the library at Oxford University was to receive free copies of all new books printed by members of the Company (Gilchrist, 2005, op. cit. p. 177).

The National Library of Scotland is rooted back to 1689 when established as the Library of the Faculty of Advocates – the Library of Lawyers of Scotland. It gained the right of legal deposit in 1710. In 1925 the Faculty of Advocates donated their library to the nation and the National Library has been founded (Wade, 2007). An Act of Parliament that brought together eight institutions including the British Museum and the National Lending Library for Science and Technology established the British Library in 1972. It is the official institution for producing and publishing the British National Bibliography (BNB), (The British Library’s strategy, 2005-2008, op. cit.).

As discussed in the Web archiving initiatives in section (1.9.1.1) in Chapter (1) regarding the extension of the legislation to cover non-print materials including Web-based content, the Library adopts a voluntary approach for preservation of Web-based content until the necessary regulations are passed (Tuck, 2004, op. cit.). The Library relies on the 2003 Legal Deposit Libraries Act and the 1963 Copyright Act, which obliges publishers and distributors in the UK and Ireland to deposit their publications. The scope of the Act also covers digital publications. In Ireland, the Copyright and Related Rights Act 2000 proposes to extend
legal deposit to digital formats (The British Library, 2006).

In Iceland, The National e-Library has reviewed its deposit legislation in 2002 and from 2003 it has begun harvesting the Icelandic part of the Web-based content. This is successfully carried out and forms the most recent part of the national e-library (Hannesdottir, 2005,p.160).

The Legal Deposit Advisory Panel (LDAP) re-launched in March 2007 the voluntary scheme for the deposit of offline digital and microform publications with the UK Legal Deposit Libraries. An e-journals pilot scheme to test the technical infrastructure, mechanisms and procedures relating to the deposit, ingest, storage and preservation of e-journals has been carried out by the Library under the auspices of the Joint Committee on Legal Deposit (JCLD) working group on e-journals. Volunteer participants were sought through UK publisher trade organisations at the end of 2004 and 23 publishers agreed to deposit over 200 journal titles, offering a diversity of subjects and complexity of formats. A voluntary deposit scheme for e-journals, developed by a joint JCLD and LDAP group, began in June 2007 (The British Library Website).
2.6.4 Die Deutsche Bibliothek

Historically, there are three libraries in Germany working as national libraries as follows:

- The German Library in Frankfurt and Leipzig;
- The Berlin State Library – Prussian Cultural Heritage;
- The Bavarian State Library in Munich (Botte, 2003).

The German National Library is established as Deutsche Bucherei in 1912 by publishers in Leipzig, and, again as Deutsche Bibliothek (the German National Library) after the Second World War in 1945, in Frankfurt, then established an archive for music in West Berlin, stands for national legal deposit collection two copies of each of 90,000 items published in Germany every year. The Library also receives deposit collections from the German-speaking part of Austria and Switzerland and collects German language materials worldwide. For many years it also collects digital library materials including Web-based content and publishes the German National Bibliography in both analogue and digital formats. Munich and Berlin libraries owe their reputation to their precious manuscripts, early imprints and non conservable items such as the Gutenberg Bible or original hand-written musical scores of Bach and the Beethoven ninth symphony manuscript (Lux, 2003).

Under the 1969 ‘Gesetz über die Deutsche Bibliothek’ (German Library Act), German publishers are required to deposit two copies of print, microform and physical format e-publications in the Library. The regime has
been extended to sound recordings in 1973. Guidelines for voluntary deposit of online content under the auspices of the Task Group on the e-Deposit Library were issued in 2000. Those guidelines currently rely on item-by-item agreements (Ibid.).

2.6.5 The BnF

The French revised legal deposit of 1992 has come into force on 23 June 1994. It requires legal deposit of printed, graphic, photographic, sound, audiovisual and multimedia documents, whatever the technical means of production, as soon as they are made accessible to the public by the publication of a physical carrier (France Legal Deposit, 1992).

The legislation does not cover online electronic publications. Deposit of CD-ROMs has been enforced since 1994. Access policies restrict networking of those items. France’s legal deposit advisory body, the 'Conseil scientifique du dépôt légal’ (CSDL) (Scientific Council for Legal Deposit), recommended in 2000 that the legislation be extended to include online content (BnF, 2006).

The initiative of the BnF’s for Web archiving derives from the French legislation and tradition regarding legal deposit and the obligations. Together with other institutions, the BnF is in charge of the legal deposit of all materials published in the country since its establishment as a legal deposit institution in 1537. French Publishers, for example, have the duty to deposit copies of all published material in the Library, which is responsible for preserving them with no limitation of time. The BnF also has the duty to process its preserved items by the Agence Bibliographique Nationale (National
Bibliography Agency), which publishes the Bibliographie nationale Françoise (National French Bibliography). Access to preserved materials is restricted to be used inside the Library building and for researchers only (Ibid).

The scope of the new legal deposit law in France is extended to cover the Web. On August 3rd, 2006 the French Parliament approved a legislation, which entitles the BnF rights and obligations concerning the legal deposit of the Web-based content. The extension of the legislation to include the Web is a reality with the vote followed by the official publication of the DADVSI\textsuperscript{56} law. The legislation is criticised because of the nature regarding connection to the national territory, and thus the exact scope of the clause ‘electronic publications’ still remain ambiguous and needs to be clarified in the decree which published as supplement in 2007 (Ibid.).

The National Audiovisual Institute (INA) has been held responsible for radio and television legal deposit since 1992 though enforced in 1995. As part of extended mission, the Institute is entitled to collect and preserve the French Web-based content of media and audiovisual importance while the BnF is responsible for preserving writings and newspapers in particular and other Web-based content generally (Drugeon, 2005).

\textsuperscript{56} Stands for Droit d'auteur et droits voisins dans la société de l'information (Law on Authors’ Rights and Related Rights in the Information Society).
2.6.6 The Royal Library of Denmark

Denmark’s Act on Copyright Deposit of Published Works, in effect from 1998, replaced the 1927 Copyright Deposit of Printed Matter to Public Libraries Act. The depository institution is the Royal Library of Denmark. It aims to cover all published material (including physical format electronic publications and static Internet content) but it does not cover dynamic online e-documents. Archived Web content is viewable only on stand-alone devices in the Library and cannot be copied by users (Henriksen, 1999).

A new Legal Deposit law, authorising harvesting of the Web, has been passed in December 2004, with effect from 1 July 2005. Article (3) covers ‘materials published in electronic communications network’, including works published in physical form, regardless of medium, material made public via electronic communication networks, authorising harvesting of materials made public on ‘.dk’ domains, the National Library of Denmark, already harvested more than 34 terabytes from the Web (Jacobsen et al. 2007).

2.6.7 Sweden, Iceland, Finland and Norway

Sweden’s 1993 Legal Deposit Act covers print, audio-visual and physical format electronic publications. An earlier Legal Deposit Act, in 1978, also established the National Archive of Recorded Sound and Moving Images. Amendments to extend the 1993 coverage to online content have not progressed. A special decree of 2002 authorises the Royal Library of Sweden to collect Swedish sites on the Web, with public access to that content on its
premises. The decree reflects the library’s Kulturarw3 project, launched in 1996, centred on automated harvesting of those sites (Sweden Legal Deposit, 1993).

Finland’s 1980 Legal Deposit Act covers printed and visual material. It was extended in 1981 to cover sound recordings. A 2000 report on revision of the Act called for extension to cover the legal deposit of physical format and online electronic publications. The latter includes works considered to be ‘true’ electronic publications (e.g. e-books and newspapers) - subject to a formal deposit arrangement - and content that would not involve deposit action by the publisher (i.e. would be collected automatically). Printers and publishers are currently required to provide with six copies of books and journals, two copies of newspapers and microforms, and two copies of ‘audio and visual recordings’. Repositories include the Helsinki University Library and the Turku University Library (Finland Legal Deposit, 1980).

A separate 1984 Act on Archives for Motion pictures covers Legal Deposit of motion pictures, films and videos in the Finnish Film Archive (Finland, 1984).

In Norway the wide-ranging 1989 Legal Deposit Act, in effect from 1990, covers books and other printed materials, sound recordings, films and videos, and some digital publications. Publishers are generally required to provide seven copies of the publication. Physical format electronic publications and static Web-based content is covered by the Act through inclusion of any works ‘that can be read, heard, broadcast or transmitted’ (Norway Legal Deposit, 1989).
2.6.8 The Netherlands and Belgium

The Netherlands, as one of the few countries in the world without legal deposit legislation, relies on voluntary arrangements. Deposit is based on individual agreements with publishers under the auspices of the national publishers association. Most Dutch printed material is deposited with the 'Koninklijke Bibliotheek' (KB), (Dutch National Library). In 1996 a voluntary deposit scheme for physical format electronic publications such as CD-ROMs and magnetic disks commenced (Belgium Legal Deposit, 1965).

A 1999 general arrangement for voluntary deposit of electronic publications extended that scheme to cover online and offline materials. Dynamic content is in principle covered by the agreement but will not be sought until technical problems are fully addressed. Access to electronic deposited publications is available from the KB premises only. Belgium’s Legal Deposit law dates from 1965 and is essentially concerned with printed works. Amendment is currently under consideration. The law provides for deposit of a single copy in the ‘Koninklijke Bibliotheek van Belgie’, the federal library (Ibid.).

2.6.9 The National Diet Library

Before the Second World War the Japan legal deposit law was totally different from what it is today. During this period there were few laws and regulations relating to legal deposit, including the Publishing Law and the Newspaper Law. The main purpose of these laws is intended to curb publishing for the purpose of controlling public order and morals. Under these laws, publishers have been obligated to
deposit two copies of their publications with the Interior Ministry, one of which has been sent to the Imperial Library (one of the predecessors of the National Diet Library (NDL), (NDL, 2006).

It is important to mention that on October 1st 1999 Japan has enacted two copyright laws in compliance with the WIPO Treaties. First, the Copyright Law of Japan concerning the circumvention of technological copyright protection measures targeting coping in particular. Second, the Japanese law on unfair competition, which relates to the TPMs controlling access to the protected works. Unlike the US law, Japanese law does not regard the act of circumvention per se as illegal. Furthermore, the laws do not provide any exceptions to the TPMs, which is explained by the fact that the act of circumventing these measures is not prohibited (Lepage, 2002, op. cit.).

The NDL is founded in 1948 under the direction of the US Library Mission. Modelled after the Library of Congress. It has two purposes, which are to assist the legislative activities of the Library as a parliamentary library and to provide library service for the Japanese public as a national library. Its legal deposit law mandates that copies of all new publications published in Japan have to be sent to its repository in accordance with Article (24) and (25) of the NDL Law (National Diet Library Law, 1948). The new legal deposit system has been in many points designed after the US Code, especially in terms of legal deposit of government publications (The National Diet Library, 2006, op. Cit.).

The legal deposit law covers publications include books, pamphlets, serials, musical scores, maps, charts films and phonographic
records and texts, images, sounds, or programmes recorded by electronic, magnetic, or other methods which cannot be directly perceived by human senses’ to deposit a copy in the parliamentary library (Yokoyama, 2006).

The Legal Deposit System Council is established on April 1, 1999 following the reorganisation of the former Legal Deposit System Research Council. The purpose of the Council is to contribute to the improvement and proper management of the legal deposit law. It has two main activities:

- to investigate important issues on the legal deposit law and the amount of compensation at the request of the Chief Librarian; and
- to make recommendations to the Chief Librarian concerning the above-mentioned issues (The National Diet Library, 2006, op. cit.).

Thus later in 2002 the Library consulted the Legal Deposit System Council to seek their views on the legal deposit of networked electronic publications. The Council submitted a report in 2004, which made it clear, following the conclusion of the Legal Deposit System Research Council (LDSRC), that incorporation of networked electronic publications into the legal deposit system is inappropriate in the light of the fundamental principles of legal deposit system – publications should reach the library, completeness of coverage; and imposing obligatory submission of their publications on publishers – and because of the characteristics of networked electronic publications. The Council suggested other methods of collecting such networked electronic publications:
coverage and means of acquisition, and ways to look at issues relating to copyright and compensation (Yokoyama, 2006, op. cit.).

In the 2004 report, the Library has encountered copyright problems when tried to collect and disseminate Web-based content. Implementation of the provision of Article (31) of the Copyright Law, for instance, limits the reproduction right in relation to preservation of ‘library materials’. The preservation of this content would be made mainly by backup or migration and this might exceed the conception of ‘preservation’ as stated in the Copyright Law (The Legal Deposit System Council, 2004).

The NDL has prepared legislation for the acquisition of Web resources separately from their existing legal deposit system and plans to conduct bibliographic control of Web resources (Knutsen, 2006).

2.6.10 National Library of South Korea

This Library is established in October 15, 1945 in Seoul. It disseminates their newly collected bibliographic information to users and it has been establishing and providing catalogues and database service on a daily base for materials collected through the legal deposit since 2005 (Chi-Ju, 2006).

By 2005, the Library holdings are 5,658,476 volumes managed by 228 library staff. The National Library of Children and Youth is opened in 2006 while the National Digital Library is launched in 2008 (Chung, 2006).

In its 60th Anniversary on 15th October 2005, the ‘National Library of Korea 2010’ has been announced as set of strategies and future plans for becoming a world model library. The strategy aims to achieve the following goals:
• Pride and repository of Korean knowledge heritage;
• Distribution and provision centre of national knowledge and information resources;
• Library policy and research hub;
• Global portal and gateway for library and information centres (Yoon et al. 2006).

The first Library Act in South Korea has been enacted in 1963, and the legal deposit has been introduced. Then the Act has been revised in 1987. According to Article (17) of the Library and Reading Promotion Act revised in 2003, two copies of all documents published or produced in South Korea should be submitted to the Library within 30 days from the date of publication by the publisher (company or individual), producer or central and municipal government (Chung, 2006, op. cit.).

In South Korea the legal deposit of electronic material depends on copyright regulations. Reference to the copyright law revised on 2003, it is not easy to use the electronic publications since there are many restrictions for the number of simultaneous users at the libraries and the available publications by the date of publication and its sale (Lee, 2006; Chung, 2006, op. cit.).

In 1992, Korea Electronic Publication Association (KPEA) has been established for the exchange of information and development of the electronic publications, standardisation and certification. In July 2004, KEPA has been chosen as an institution for submission (legal deposit) of the electronic publications. The number of the legal deposit materials are 4,128,694 among 5,786,393 the total number of collection in March 31, 2006. The future Act will cover static and dynamic online resources.
Acquisition, through legal deposit, accounts for approximately 70% of the overall library collections. Web-based content are not included within the scope of the deposit due to lack of legal provisions. The Library solved such problem by encouraging voluntary deposit (Lee, 2006; Chung, 2006, op. cit.).

2.6.11 National Library of Australia

This Library is responsible most of all to support the exploration and promotion of methods to preserve Web-based content collected by Web harvesting. It aims to work out a survey of existing standards, guidelines and codes for preservation of digital materials in cooperation with the Preservation and Conservation Section of IFLA. It is leading the work of the Deep Web Working Group of the IIPC and works in closer cooperation with the BnF to investigate the identification, acquisition, storage and display of publications and database driven Websites (Bellingham, 1999).

The Copyright Act of 1968, which is enacted on 4 March 2001 pursuant to the 1996 WIPO Treaties, meant to prevent copying through implementation of the TPM though offers certain exceptions to users as recognised in the Act (Lepage, 2002, op. cit.).

In Australia the Act recognises the need for preservation copying and reformatting but in a very limited way hence adding further obstacles for preservation. The Act does allow a library or archives to make copies of a limited range of original documentary materials. It, thereby, allows a replacement copy to be made pursuant to section 51A(1)(b) in regard to static library material (Australia Copyright Act, 1968).

The aim of documentary preservation in Australia is to ensure sustainable public
access to information, wherever possible through preventive preservation actions. Legal deposit legislation subsumed under section (201) of the Act does contribute to ensuring access and preservation of certain categories of published works in Australia (Bellingham, 1999, op. cit.).

The Act serves to limit access to published and unpublished materials through applying tests of fair dealing or fair use. The rationale is to encourage and motivate authors and creators to keep producing and contributing to human knowledge. While this logic makes sense, it is difficult to anticipate how this incentive can be effective for fifty years after an author’s death. In this situation, there appears to be no particular concern for the preservation of library material, but rather mere concern for economic interests (Australia Copyright Act, 1968, op. cit.).

In terms of exceptions of Web-based content for digital preservation purposes, the PANDORA Archive refrains, as stated in the literature review in Chapter (1), section (1.9.1.1.8) from implementing full coverage for the collection policy due to copyright barriers because there are no legal deposit provisions for capturing dynamic Web-based content (Philips, 2004, op. cit.).

2.6.12 National Libraries in Africa

Africa has 34 least-developed countries and 20 developing countries. They are all sovereign states, with different laws and jurisdictions. Priority in Africa is not given to library services. For most, if not all of African countries, illiteracy, unemployment, lack of infrastructure and resources, famine, disease, conflict, crippling debt, and mere day-to-day
survival are far more pressing issues than IPRs especially copyright. Until recently there has been no copyright co-operation or harmonisation in Africa. Current copyright regimes in Africa are inappropriate and fail to address the legitimate needs of education, libraries, and people with visual and auditory disabilities (Nicholson, 2006, op. cit.).

The major challenge that African communities suffer from is the digital divide phenomenon. Internet penetration rates in Africa are very low. The ITU and WSIS indicators provide that:

- In 2004, less than three out of every 100 Africans use the Internet, compared with an average of one out of every two inhabitants of the G8 countries (Canada, France, Germany, Italy, Japan, Russia, the UK and the US).
- There are more than eight times as many Internet users in the US than in the entire African continent.
- There are more than three times as many Internet users in Japan as in the entire African continent.
- There are more than twice as many Internet users in Germany than in the entire African continent.
- The entire African continent - home to over 50 countries - has fewer Internet users than France alone.
- There are more Internet users in Seoul, than all of sub-Saharan Africa, excluding South Africa.
- Switzerland, host of the WSIS Summit, has five times the Internet penetration rate of Tunisia, host of the second Summit (ITU & WSIS, 2003-2005).
The Sudan is no exception of the above equations but it is expected that the rates may grow in the coming generations as education expands and advances.

Summary

Some international concerned organisations state that the implementation of exceptions in the IPRs regime is basically the choice of a national legislature to decide. Berne Convention (1886, op. cit.) and TRIPS agreement (1994, op. cit.) simply requires that, where they exist in national law, they should not unreasonably impair the copyright owner’s interests. This is expressed in a so-called ‘three-step test’ which appears in the Berne Convention concerning exceptions that allow users to make copies without the copyright owner’s permission, and which TRIPS then applied more generally to all forms of copyright as in Article 9(2) of Berne, Article (13) of TRIPS and the US ‘Fair Use’ doctrine. In the Berne Convention this provides that countries can permit reproduction of copyrighted works without the permission of the copyright owner in certain special cases, provided that such reproduction does not conflict with a normal exploitation of the work and shall not unreasonably prejudice the legitimate interests of the author (IFLA-IFLA-CCLM, 2002, op. cit.).

In the settlement of Authors Guild v Google, Inc. case as mentioned, it is obvious that the settlement has the potential to provide unprecedented public access to digital library and Web archive services. At the same time, in the absence of competition for the services, the settlement of Google Digital Library could lead to monopoly of information and compromise the essential library values such as equity of access to information, patron privacy, and intellectual freedom. Also there is a concern that the expanded
services permitted under territorial settlement will be available only to users in the US unless the company reaches agreements with right holders in other countries, there will be an ever-widening global digital divide in access to content. The settlement also need to consider the issue of long-term preservation and access.

Article (8) of the UNESCO Charter, as mentioned in Chapter (1), section (1.9.1.1.), emphasises that the UNESCO’s member states should protect their digital heritage through inaction of proper legal and institutional frameworks. This legal framework should embrace national preservation policy, archive legislation and legal or voluntary deposit of this heritage in libraries, archives, museums and other public repositories (UNESCO, 2003, op. cit., p.64).

National Supreme Court and other federal courts are responsible for handling dispute settlement and arbitration regarding cases of IPRs infringements (WIPO, 2003).

All in all, the theoretical and conceptual framework elaborated the problems of the study in this Chapter is ended. The next Chapter begins the empirical investigation and provides the methodology with the aim to find out solution of these problems. This is the end of the theoretical part of the study and the empirical part starts with the next Chapter.
CHAPTER 3
METHODOLOGY

Overview

The goal of this empirical part of the study is to describe the quantitative procedures and measure for identifying the impacts of IPRs on preservation of Web-based content as discussed in Chapters (1) and (2) respectively. The research investigates the Sudan National Library’s legal deposit readiness for preservation of such content. This Chapter is consisted of two parts: an action and survey methodologies, in which some national libraries and archives of countries that maintain live Websites were studied, as they held responsible for preservation of Web content. This Chapter aims to describe the research design, the procedures that are followed, the population of the study, the instruments that are developed for data collection, and the methods for data analysis. It begins by the establishment of the context of the study through presentation of the rationale of the study.
3.1 Rationale for the study

Research objectives as Thomas and Brubaker (2007) demonstrate are divided into two main purposes for both the master’s degree and doctoral projects as follows:

- to provide graduate students guided practice in conducting and presenting research; and
- to make a contribution to the world’s fund of knowledge or to improve the conduct of some activity.

Reference to the above objectives, this research is to discover new facts; information and data intended to solve the problem of the study through building the Sudan Web Archive. To put the study in a context, the next part provides a convenient research methodology.

Researches in measuring the Web size are rare. This is attributed to the vastness and dynamic nature of the Web. These challenges made Snelson (2005) to say that the process of finding solutions to these challenges may involve revision of traditional research methodologies. Consider the basic research procedure of defining a population and selecting a suitable sample to study. In Web-based research, complex problems emerge when defining populations or selecting samples. For instance, if the population is defined to be all existing Web pages, then access problems are encountered. A portion of the total population of all existing Web pages is private, and public access to commercial content is restricted. If the private pages are eliminated from consideration for research, then the population under study may be redefined as the set of Web pages open to the general public.

To overcome such problems, the Web Characterisation Project tries to fill a gap in Web
research, which has tended to neglect analysis of the actual content of the Web or authoritative descriptions of its size and information structure. In short, the Project focuses on description and analysis of the Web as an information collection. It also helps librarians to integrate Web resources into their holdings. The project is concerned with the development of new metrics for Web description (Covert, 2000, p.14).

Research in LIS started with articles by Jarvelin and Vakkari (1990, 1993) on trends on international LIS research. They described such trends when compared distribution of topics, approaches and methods in the years 1965, 1975 and 1985.

This research is situated within the approach of the Webometrics studies as referred to in Chapter (1), section (1.9.1). In that section, the intended aim is to provide logical background from general to specific while this part intends to link it to the major trends in research methods and those who contributed to its emergence. The Webometrics field, according to Aguillo (2002), is still in its infancy as a scientific method with its own various theories to be built, tasks to be done, units to be defined, methods to be developed and problems to be solved.

It is concerned with the quantitative aspects of how different types of information are generated, organised, distributed and utilised by different users in different contexts. This development arose during the first half of the twentieth century from statistical studies of bibliographies and scientific journals (Hertzel, 1987).

The development of online citation analysis parallels the later avalanche of Webometrics studies enabled by access to large-scale Web-based content. In particular, the apparent resemblance between citation networks and the hypertextual inter-document structures of the Web triggered much interest from the last decade of the twentieth
century. Further, the central bibliometric measures of co-citation and bibliographic coupling have been applied to studies of Web clustering, growth and searching (Larson, 1996).

As referred to earlier, Webometrics thus offers potentials for tracking aspects of scientific endeavour traditionally more hidden from bibliometric or scientometric studies, such as the use of research results in teaching and by the general public and the actual use of scientific Web pages (Bjorneborn & Ingwersen, 2001; Cronin, 2001; Thelwall & Wilkinson, 2003a).

A range of new terminologies for the emerging research field have been rapidly proposed from the mid-1990s, for instance, ‘netometrics’ (Bossy, 1995); ‘Webometry’ (Abraham, 1996); ‘Internetometrics’ (Almind & Ingwersen, 1996); ‘Webometrics’ (Almind & Ingwersen, 1997); ‘cybermetrics’ (journal started 1997 by Isidro Aguillo); and Web ‘bibliometry’ (Chakrabarti et al. 2002).

The breadth of coverage of cybermetrics and Webometrics implies large overlaps with proliferating computer-science-based approaches in analyses of Web-based content, link structures, Web usage, and Web technologies. A range of such approaches has emerged since the mid-1990s with phrases such as ‘Web Mining’ (Etzioni, 1996; Cooley et al. 1997; Kosala & Blockeel, 2000), ‘Web Ecology’ (Pitkow, 1997; Chi et al. 1998; Huberman, 2001), ‘Cyber Geography’ or ‘Cyber Cartography’ (Girardin, 1995, 1996; Dodge, 1999; Dodge & Kitchin, 2001), ‘Web Graph Analysis’ (Kleinberg et al., 1999; Broder et al., 2000), and ‘Web Dynamics’ (Levene & Poulouvassilis, 2004).

The reason for using the term ‘Webometrics’ in this context could be to denote a close lineage to ‘bibliometrics’ and ‘informetrics’ and stress a LIS perspective on Web studies as noted above. In this context, the earlier mentioned term ‘bibliometry’,
as used by Chakrabarti et al. (2002, op. cit.), is especially interesting because computer scientists thus recognise the heritage in bibliometric research to be drawn upon in Web studies. Other computer science approaches to link structure analysis also pay tribute to inspiration from citation studies (Pitkow & Pirolli, 1997; Kleinberg, 1999; Kosala & Blockeel, 2000, op cit.).

Based on Ingwersen (2006) assumption, from the dawn of the twenty-first century, the Webometrics field has expanded into various studies including the following:

- Web Indicators – many laboratory groups around the world work to reinforce the quality of such measures and correlation studies;
- Web space studies – observation of the relationship to social networks and studies of other social phenomena of the Web;
- Web data collection – this area is of great importance of the Webometrics, both in terms of quality assessments of search engines but also concerned with how to apply Web crawlers, adequate sampling methods, limitations as to Webometric analyses, and so on. There is, as in other Informetric analysis work, a strong link to the field of information retrieval (IR) – as of Web information retrieval and Web data mining;
- Web link analyses – this is a particular well defined emerging scientific domain.

A central difference from traditional scientific databases and archives is the dynamics of the Web. Time plays a crucial role on the Web.
3.2 The Action Research Method

In order to solve the problems of this study, the action research methodology is employed to discover a solution for the disappearance of Web-based content through envisioning certain strategies for building the Sudan Web Archive to capture this rapid and dynamic content. This research inquiry has potential applications for addressing issues broadly related to acquiring, managing and using information. The vast majority of LIS research is applied or action research (McClure & Bishop, 1989) in nature as opposed to basic research. Action research identifies problems in specific setting and suggests strategies to deal with those problems (Busha & Harter, 1980; Hernon & McCulre, 1990; Powell, 1997).

Therefore, the aim of adopting the action research method herein is not only to find a solution for the problems of this study, but also, as McClure (1991) argues, to take greater responsibility for increasing the effectiveness with which research is communicated to the LIS profession, especially those in decision-making positions concerning preservation of Web-based content before disappearing forever.

3.3 The Survey Research Method

The survey method is often viewed as a relatively popular, easy to design, administer, quick and cheap method of obtaining data from a large number of respondents.

The wide availability of powerful desktop computers and statistical software has made them easier to tabulate, manipulate and interpret. It is possible to use interactive survey techniques, with the responses immediately entered into a database and processed (Gliner & Morgan, 2000).

The costs of Web and e-mail surveys are generally lower than for data collected by interview,
especially if the population is widespread. LIS researchers (Rochester & Vakkari, 1997) investigated three most popular research methods used universally. In the international literature the conceptual research method remained the most popular over time at 29% in 1965 and 1975, 23% in 1985. Also the survey method is often used at 23% in 1965, 20% in 1975 and 23% in 1985. Historical method is used at 11% in 1965, and replaced in popularity by system design in 1975 and 1985, at 15%. The survey is the second most popular in the first two periods, and in 1985 and 1995 has been most popular at 20% and 29% respectively. Survey and conceptual methods are the only ones also popular in the international literature.

As long as Drake (2003) adds, the survey (questionnaires and interviews) method is rated top during the period from 1975 to 1979 at 163, and from 1985 to 1989 at 149, and from 1990 to 1994 at 118 in the library science dissertations. But it is ranked second during the period from 1975 to 1979 at 18, and from 1985 to 1989 at 40, and from 1990 to 1994 at 48 in information science dissertations.

This survey method is an exploratory quantitative study. It uses a cross-sectional survey i.e. a survey in which the data is collected at just certain point in time to find out how other custodial institutions deal with barriers of IPRs in preservation of their Web domains. The design of the questionnaire scales are derived from the dependent variables discussed in the literature review in sections (1.9.3 and 1.9.4). These variables focused on the importance of preservation of Web resources, current legal status of preservation of Web resources, public domain information, threats to storage of Web content, preservation budgets, responsibility, file formats and harvesting systems. It also provided space for comments and feedback.
3.3.1 Data Collection Methods

Quantitative data about the respondents were collected, including names, positions, and type of institutions (See Appendix (A) Table No. (5.1) and Figure No. (5.2)). Possible relationships among the demographic variables and importance of Web resources, legal readiness, technical and financial problems have been explored. The survey questionnaire has been timed for a limited period of time (three months).

The responses are received via e-mails and only one via a fax. Follow-up reminders are made to increase the response rates. Interviews also conducted as discussed below. These combination of data collection methods are used to collect and analyse data on the nature and types of preservation of Web-based content in the national libraries and archives, the policies and programmes in place for preservation, preservation needs and requirements, and future plans for developing Web-based preservation programmes.

3.3.2 Specific Procedures

The respondents needed to be familiar and regular e-mail users to respond to the questionnaire attachments. New logbook of folders or files is created on 'My Documents' for data storage. Each folder or file is labelled according to the responders’ national library or archive. Variables have been classified, encoded and saved on these folders and files. Data located in the returns were encoded and subjected to descriptive statistical techniques using Excell spreadsheets and SPSS\textsuperscript{57} system. Analysis of variance (ANOVA), Pearson correlation, ratio statistics, regression, reliability, and T-Test (see appendix

\textsuperscript{57}Statistical Package for Social Science.
A) are calculated as measurements for enabling generalisations to be made. A free version of SPSS version 16.0 software is downloaded from the Web in order to assist in undergoing some statistical operations. Access to the software is permitted upon registration and long in for duration of two weeks then expires. The data are migrated from SPSS worksheets into pdf formats in order to avoid software expiry before presentation of the calculations is made. The result of tables and figures are made available in the appendix (A) at the end of this study to avoid interrupting the readers’ attention.

3.3.3 Sample Population

The sample has been drawn, as mentioned earlier, from national custodial institutions in order to learn from their experience and practices in preservation of Web-based content. The approach has the advantage of gathering data from a variety of institutions with or without digital preservation responsibilities, including national libraries and archives.

However, random sample selection is not possible in this situation, since only responders willing to take the time and respond to the questionnaire did so. This may have introduced a bias and limited the sample’s representativeness and thus the generalisability of the findings. Though this limitation, the results helped identify the strategies used to establish the Sudan Web Archive (cf. Chapter 4) and provided useful data about actual Web preservation practices, types of preservation technologies, and legal features relied on. The findings could serve as a base for future experimental Webometrical research.
3.3.4 Instrumentation

Since no research with this focus had been conducted regarding preservation of Web-based content before, an instrument has been custom developed for this study, tapping preservation characteristics, strategies, and legal readiness. The questionnaire is made up of closed-ended questions mainly, but where relevant, it allows for additions and comments. The introductory section is devoted to demographic data and particulars about the type of institutions investigated.

In developing the questionnaire, care has been taken to avoid excessive length that could discourage respondents from completing it. However, the questionnaire has still been in three pages, due to the number of issues (variables) under investigation. Its questions were placed before answer’s boxes to increase usability and ensure an aesthetic layout and automatic data collection. The instrument has been pre-tested with a small sample of users (at GHD Global staff worldwide) to discover ambiguities or other problems. Suggestions were incorporated and ambiguities solved.

3.3.5 Validity Analysis

The questionnaire has been given to groups of judges for content validation. The groups were made up of experts in quality assurance and marketing as well as the supervisors of this study. A GHD Global quality assurance administrator with more than 25 years of quality assurance experience who has done researches, in coordination with the researcher, comparing engineer’s reliance on building codes and standards and quality of designs who has distributed extensive surveys on the subject of
marketing for identifying the relationships between the architect’s presentation skills and business acumen. The supervisors of this study also have been informed for approval and distribution. The groups are kept aware with a statement of the purpose of the study and a description of the subjects, as well as the aims of the study. They have been asked to judge both content and format of the questions. The questionnaire is revised according to the judges’ suggestions and resubmitted for final evaluation. The judges confirmed that the questionnaire tapped the data needed to answer the research questions (i.e. validity) and used appropriate language that is unambiguous.

3.3.6 Reliability Analysis

A factor analysis is performed so as to determine if various variables could be grouped into factors (see tables and figures in appendix A). The groups are divided into three classes encoded as (QuestionB, QuestionC, and QuestionD) so that SPSS software could deal with them easily. The consistency or reliability of specific scales in the questionnaire is checked to reflect the construct it measures through test process using the above-mentioned software to compute the Cronbach alpha coefficient. The items are revealed as strongly related to alpha reliability coefficient of the index at 0.8 level (see appendix (A), Figure 5.5).
3.3.7 Response Rate and Survey Population

The questionnaires have been forwarded to 146 national library and archives. There are 857 of them returned from national libraries in North America, the Caribbean, Europe, Asia and Australia in the first package. The second package received 343 responses after conducting continuous reminders (total of 1200). Based on formal classification, the responses are tabulated by institution as in Figure (5.1) in appendix (A).

3.3.8 Size of Sample and Response Bias

The potential bias in response is toward institutions with some digital materials in their holdings in order to compare the characteristics of non-respondents with institutions that responded; it is possible that institutions, which consider digital preservation an important problem because they have digital materials in their holdings, have been more motivated to respond to the survey. This cautionary note is reinforced by anecdotal data collected from respondents while they have been completing the questionnaires and through informal feedback received after they have been completed. Several institutions reported ease of completing the survey and requested a copy of the results. Compiling all of such data and answering the questionnaire thoroughly required a significant amount of time on the part of the staff of the institutions, suggesting that a high level of motivation is necessary for them to respond.

Findings reached through the survey and interview tools are cohesively connected with the investigations carried out throughout this Chapter and the Chapters to come in order to meet the research objectives stated in Chapter (1).


3.3.9 Formats for Result Presentation

The data have been analysed to identify the respondents’ Web-based preservation plans and practices for different national domain spaces, liability, responsibility, funding, and threats. Descriptive statistics used in order to measure variables for extracting any potential relationships. The data are presented through tables and pie charts of variables and figures as shown in appendix (A) at the end of the study.

3.4 Interviews

The semi-structured interviews are conducted to verify how the Sudan National Library and National Record Office deal with preservation of Web-based content and whether they would adopt an initiative for such purpose. A semi-structured interview is an oral administration of a questionnaire, and the data obtained to serve for comparative purposes. The interview agenda targeted aspects of the legal deposit readiness that could pave the way for preservation of the national Web space. In addition, questions similar to those in the e-mail questionnaire were included for comprehensiveness purposes (See Appendix A). Unlike traditional ethnographic observation, in which researchers immerse themselves in the situation for weeks or months, the process is limited to a short scheduled one-day period.

Summary

The next Chapter proposes establishment of Web legal deposit to enable achieving the goals of the study by building the Sudan Web Archive.
CHAPTER 4
BUILDING THE SUDAN WEB ARCHIVE

Overview

As discussed in Chapter (3), the action research methodology is employed in order to find out solution for the problems of the study. The successful experience on Web archiving initiatives discussed in the literature review in section (1.9.1.1) as well as the results of the survey (see analysis of findings in Chapter 5), support establishment of the Sudan Web Archive in order to solve the problems of the ephemeral nature of the disappearing Web-based content. The objectives of this Chapter are to provide some Internet demographic statistics about the country, the preservation programmes, IPRs administration, the Web legal deposit proposal, the preservation strategies and the plan for building the proposed Web Archive.
4.1 Internet Demography

The Internet World Stats (2008) provides that African Internet users estimated at 3.6% of 96.4% of the world in 2008. In the Sudan, user growth for the period from 2000 to 2008 is at 4,900.0% and that latest data of users in the country are 1,500,000 (1.5%) among top ten African countries. Statistics of the World Bank (2008, p.92) show that the populations of the Sudan are 36.2 million in 2007. The literacy rate is 61% in the 2006 World Bank report. These statistics illustrate, from the researcher’s point of view, that 39% of the populations are illiterate, i.e. unable to get use of the traditional print media hence digitally excluded.

The story of the Internet service in the country goes back to 1997. The ‘.sd’ registry has been first delegated by the IANA on 6th March 1997 to Ihab I. Osman of Sudan OnLine, Inc. By mid-2001 the ‘.sd’ registry is essentially dead. On 18th November 2002 the Sudan Internet Society (SIS) has been re-delegated the registry (IANA, 2002). The researcher is concerned that the duration from the death and resurrection of the ‘.sd’ registry showed above might have caused loss of great national Web-based content. Internet users in 2000 were 1 person per 1,000 then the number grew to 77 persons per 1,000 in 2007 (The World Bank, 2008, op. cit., p.93). However, it is vital to notice that most of the Internet users are concentrated in the Khartoum State while other States still lagging behind. The rate of digital divide is so high within rural areas.

An OECD\textsuperscript{58} Report (2004, op. cit.) confirms that improving ICT access and the quality of use is an important issue. Computers may be spreading, but not all homes, schools, or businesses have taken them up or are using them as they might. This is

\textsuperscript{58} Organisation for Economic Co-operation and Development.
added to the linguistic problem, which applies to many of the Sudanese communities specially those who neither speak Arabic nor English for they face serious communicative barriers.

Before 1994, the public sector has monopolised the telecommunications services but the situation has dramatically changed since the establishment of the National Telecommunication Corporation (NTC) which aimed at opening up the telecommunication services for the private sector in order to develop the telecommunications services in the country and transform the Wire and Wireless Communication Department owned by the government into a new company subject to instruments of supply and demand known as the Sudan Telecommunications Company (Sudatel). According to the Arabic Network for Human Rights Information (HRINFO, 2006) Sudatel has established new telecommunications network from scratch. Another pro-government company has emerged known as ‘Sudanet’ aiming to run the Internet services on a monopoly basis in 1998.

In 1996 the first ISP, the Sudanese Internet Service Company (Sudanet) has introduced its service in the country. In June 1998 an agreement has been signed between Sudatel and Sudanet and accordingly Sudanet became the only ISP provider in the Sudan. There are only 14 cities have been yet connected (El Tigani, n.d.).

Sudanet remained as the only ISP until 2005 when monopoly is ended up in an endeavour to promote and reduce the telecommunication costs as well as bridging the gap between urban and rural areas. As a result, there are new ISPs working so far such as ‘ZinaNet’, and ‘Sudacom’ (HRINFO, 2006, Op. cit.). However, it is worth to say that the SIS is the local registry for the ccTLD name ‘.sd’ in the Sudan (SIS Website, op. cit.).

It is reported that most users of the Internet Cafes in the country are students 30%. Generally, majority of the purposes of Internet navigation
ranges from studying and entertainment 30%, research 15%, and e-mail 40%. The country has the highest rate of female Internet users of 40%, which is growing compared with their counterparts in other Arab countries (HRINFO, 2006, op. cit.). Average time of Internet usage is one to two hours. Studies (Ahmed & Yousif, 2007) also claim that the Internet penetration among the Sudanese doctors is low and the matter is worse for those in other professions in rural areas. HRINFO (2006, op. cit.) provides that the NTC has established a sponsorship unit for Internet surveillance and filtering to protect users from harmful content such as pornography, antireligious and other material, it is reported that the volume of the pornographic content blocked is more than 95% of the total blocked content over the '.sd' domain.

4.2 Preservation Programmes in the Sudan

In the Sudan, there are more than three institutions with responsibility of legal deposit functions. The first institution that carried out the mission of national library has been the Sudan Library at the University of Khartoum, which was called Sir Newbold Library. Sir Newbold was the British Administrative Secretary in the Sudan. The Sudan Library also held national preservation responsibility as a national library. It implements the Legal Deposit Act of 1971.59

The second institution is the Documentation Centre of the National Research Centre (DCNRC) at the Ministry of Higher Education had also established a library specialised in documenting researches and studies and held accountable for current researches register, national research

---

59 دار الوثائق القومية, (د.ت.), كتب تعريفي عن دار الوثائق ص.ص. 2-3. الخرطوم, القومية.
bibliography and publishes the Sudan Science Abstract Journal (Karar, 2005).

The third institution is the Sudan National Electronic Library (SNEL) which was established in the 1990s by the Sudan Telecommunications Company (SUDATEL) as an ambitious project and is budgeted at USD$575,000 (UNESCO, 2002), and the Sudan National Library which has been launched recently after its law that have been endorsed in 1999 by the Ministry of Culture and Information with a law establishes its duties and obligations (Karar, 2005, op. cit.).

The fourth institution is the National Record Office (NRO), the oldest archival authority and reports to the Council of Ministers. The idea of the NRO began in 1916 when the British Administration established an office with responsibility to collect and organise records, correspondence and financial transactions. By the twentieth century, the office’s mission has been extended to include judicial correspondence. In 1948, the British Administration established the Sudan Archives Office, which was responsible for collecting all kinds of official documents, publications, and manuscripts of families and religious leaders in the country. In 1955, the Archives Office has been nationalised and its name was changed in 1965 to become the National Central Records Office (NCRO) then changed again under a law issued in 1982 to be renamed the NRO as a sovereign entity supervised by the Ministry of Cabinet.

According to Osman (1990, p.13) the management of archives has commenced in the Sudan since 1916 during the Condominium rule. The Central Records Office (CRO) has been established in Khartoum in 1955 under the custody of the Ministry of Interior. The first archivist who held the position officially was Professor Peter M. Holt, the
Historian of the Sudan. Late professor Mohammed Ibrahim Abu Saleem was the first Sudanese to occupy the position after Professor Holt.

The first CRO legislation has taken place after enactment of the Directive of 1953, which followed by the Central Records Act in 1965, and its general regulations in 1966. The same year witnessed approval of the Legal Deposit Law, which was revised in 1971 and obliged all the Sudanese publishers inside and outside the country to deposit a copy of their publications in the CRO, University of Khartoum Library and the Central Library of Omdurman Islamic University. The Act was revised and replaced by the 1982 Act after enabling certain amendments to form a national council in order to take care of the CRO’s affairs in addition to establishment of general policies for archives and libraries at the national level. The National Records Authority (NRA) was established pursuant to the 1982 Act. However, general regulations for the same Act were issued in 1984 (Osman, 1990, p.14, op.cit.).

It was discovered that there were weaknesses and defects in some parts of the Act and this problem led the NRA’s Board on 14th December 2000 to propose amendment of the 1982 Act to correct such defects. Reference to Article 109(1) of the 2005 Transitional Constitution of the Sudan, the President of the Republic issued an Interim Decree of the National Records Office 2005 Act.

The Act also does not include any provisions for preservation of digital materials. This may be attributed to unawareness of their value as cultural heritage as discussed earlier in Chapter (1), section (1.1.2). Article (5) of the 2005 NRO’s Act identifies the objectives of the NRO as follows:

---

61 دار الوثائق القومية. نفس المصدر السابق ص. 4
• Protects, preserves, and provides access to the national documentary heritage;
• Develops and implements advanced methods and technologies for preserving and retrieving records deposited by the governmental departments;
• Raises awareness of the importance of records as information sources and components of the Sudanese heritage;
• Displays originality of the Sudanese identity through deepening the nations' believes through their heritage that emerges from their documentary heritage\textsuperscript{63}.

To achieve the objectives as stated in Article (5) and (6) of the 2005 NRO's Act, the NRO shall practice the following functions and powers:

• preserves the national records which produced in various departments including private documents referred to in Article 17(1);
• collects and preserves national records of scientific and legal value;
• establishes rules and regulations to maintain persistence of records in all departments in order to protect literary, scientific, historical, artistic, cultural and political heritage;
• facilitates dissemination and access to records to shade light on civilization of the Sudan and implant sprit of respect and dignity;
• raises awareness of knowledge and national records;

\textsuperscript{63} \textit{Id.}
• promotes the documentary work and the associated works such as documentation, libraries and heritage;
• establishes branches for records and documentation in the vocational institutions in the centre and the states;
• establishes specialised information network for the national documentation system as a branch of the National Information Network in coordination with the National Information Centre;
• protects the public or private records related to public good and preserves them from destruction and illegal disposal;
• inspects records of the governmental departments and collects dead valuable records which have no administrative value including records of the dissolved or cancelled departments which changed their preservation functions;
• approves operations of disposal of official records which have no administrative value pursuant to rules and regulations;
• collects fees in accordance with the 1997 Procedural, Financial and Accountancy Act.

There are other institutions also involved in some way or another in advancing digital preservation in the country. Academic libraries at universities with postgraduate programmes also preserve print and digital copies of theses and dissertations though this should be the National library’s responsibility. The National Information Centre (NIC), for instance, has been established in 1999 in Khartoum to establish the National Information Network (NIN)⁶⁴.

---

In the 2006 NIC Plan, study of e-government readiness is going to be conducted in co-ordination with the NTC and a fund from the UN Development Programme (UNDP). It is noteworthy that one of the NIC’s objectives is to technically supervise the governmental Websites and establish the national information bank\textsuperscript{65}. The idea of the information bank is not new, in 1989, the Prime Minster at the then requested from Dr Radia Adam Mohamed to prepare study for establishing such bank\textsuperscript{66}.

However, the NIC also established the 2006 Cybercrimes Act, which takes force if crimes are committed partially or holistically inside or outside the country or if their impacts extends to affect domestic works. The Act aims to regulate uses of the Web and combat piracy on intellectual works\textsuperscript{67}.

Another institution is the National Telecommunication Corporation (NTC), which has been established under the 2001 Telecommunication Act in September 1996 in Khartoum. Its mission is to provide effective regulatory framework and adequate safeguards to ensure fair competition and protection of consumer interest (NTC Website, op. cit.).

The NTC is revising the 2001 Telecommunication Act in order to be more inclusive, keeping up-to-date with technical steady development in the field of telecommunications and information technology, and its blending techniques, applications and services\textsuperscript{68}. The actual responsibility of preservation of the national heritage will be discussed in the coming part.

\textsuperscript{65} http://www.nic.gov.sd

\textsuperscript{66} حمد، رضية آدم (1989). إنشاء بنك المعلومات برئاسة مجلس الوزراء. اختراع: مركز علوم المعلومات. جامعة اختراع.

\textsuperscript{67} السودان، مشروع قانون مكافحة جرائم المعلوماتية لسنة 2006. اختراع: المركز القومي للمعلومات.

\textsuperscript{68} Ibid.
4.3 The Sudan National Library

Historically, national libraries have been mandated to preserve the published cultural heritage of their countries for future generations. As discussed previously in Chapter (2), section (2.6), this mandate has usually been based on legal deposit laws that require publishers and individual authors to submit copies of all their publications to the national library. In addition to legal deposit of local publications, national libraries also serve to collect publications related to their countries from abroad in order to maintain sovereignty of information.

In prehistoric times, the practice of using a special room or building for the storage and classification of written matter in order to preserve cultural traditions and improve social organisation can be traced back to the peoples of Mesopotamia, starting with the Sumerians early in the third millennium BC or possibly earlier. The Sumerians developed a cuneiform script in which they wrote on clay tablets, not only factual records giving an insight into their daily life but also a considerable body of literature which had an enormous influence on their education and their cultural activities generally (Staikos, 2000).

Feather and Sturges introduce a definition of 'national libraries' as follows:

‘Libraries which have a responsibility to collect, maintain and preserve the nation’s literature’ (Feather & Sturges, 1997, P.1870).

There is an important distinction between national libraries and national library services. A ‘national library service’ can therefore be defined as follows:
‘an institution, primarily funded (directly or indirectly) by the state, which is responsible for providing library services of one or more kinds to communities of defined types throughout the country through a network of branch (constituent) or affiliated libraries and service points. A national library service can include a national library as one of its constituent libraries or divisions’ (UNESCO, 1997, p.7).

However, the importance of national library services depends on the quantity, quality, size and range of collections (Feather and Sturges, 1997, op. cit.). The UNESCO, states that there are three dimensions of the concept ‘national library’ paraphrased as follows:

- **Heritage**, ‘emphasis on the nation’s intellectual production, knowledge treasures.’ This type of national libraries, which emphasises this dimension, corresponds most closely to the ‘classic’ national libraries. Care of collections is the focal point.
- **Infrastructure**, ‘emphasis on national co-ordination, facilitation, leadership, and services.’ This type emphasises a more modern development. Service to the country’s libraries is the main objective.
- **Comprehensive national service**, ‘delivery of services to users, not merely in a reading room in the capital city, but throughout the country.’ This type is mainly found in developing countries. Service to the public is the central concern (UNESCO, 1997, op. cit.).
4.3.1 Functions of the National Library

According to Phillips, the primary role of a national library is to develop and maintain comprehensive collections of documentary material relating to its country and its people, to preserve them and to make them available for research now and in the future (Phillips, 2003, op. cit.). National libraries usually act as resource archives or ‘libraries of last resort’. Most national libraries provide research departments to support and disseminate the results obtained from research projects and co-ordinate ILL lending services and provide public access to their archives (Feather & Sturges, 1997, op. cit.).

The National Library of Australia Act defines the role of the national library as follows:

'developing a national collection of Library material, including a comprehensive collection of library material relating to Australia and the Australian people; making library material in the national collection available; providing library services and cooperating in library matters with authorities or persons, whether in Australia or elsewhere, concerned with library matters' (Australia National Library Act, 1960).

Though Australia follows the Anglophone legislation system but its 1960 National Library Act does not include provisions for deposit as of the UK 2003 Legal Deposit Libraries Act. The PANDORA Web Archive just deletes the proprietary content upon request of the right holders as referred to in the Literature Review in section (1.9.1.1).

Libraries in general and national libraries in particular, form an essential and irreplaceable component of the cultural, educational and
informational infrastructure of a society. Moreover, they form an irreplaceable part of the cultural heritage (Council of Europe, 2000).

The IFLA Section of National Libraries, in its medium-term programme for 1992-1997, described the nature and purpose of national libraries as follows:

'National libraries have special responsibilities, often defined in law, within a nation's library and information system. These responsibilities vary from country to another but are likely to include: the collection via legal deposit of the national imprint and its cataloguing and conservation; the provision of central services (e.g. reference, bibliographical, lending) to users both directly and through other library and information centres; the preservation and promotion of the national cultural heritage; the promotion of national cultural policy; and leadership in national literacy campaigns. National libraries often serve as a national forum for international programmes and projects. They can have a close relationship with national governments, and can act as a conduit for the views of other sectors of the profession. Occasionally they also serve the information needs of the legislature directly' (IFLA, 1992, pp.20-21).

The National Libraries Amendment Bill of 1997 of South Africa states that the functions of the National Library are:
• to build up a complete collection of published documents emanating from or relating to the country;
• to record the documents intended in the above item, to render a national bibliographic service and to act as the national bibliographic agency;
• to promote optimal access to published documents nationally and internationally;
• to provide reference and information services, nationally and internationally;
• to act as the national preservation library and to provide conservation services on a national basis;
• to reinforce awareness and appreciation of the national documentary heritage; and
• to promote information awareness and information literacy (The National Libraries Amendments Bill, 1997).

Most national libraries compile and maintain national bibliographies and operate bibliographic information centres (Feather & Sturges, 1997, op. cit.).

Categories of national library range from merely national library institutions such as the British Library - it was the British Museum - National Library of Australia, and the BnF and parliamentary libraries such as the Library of Congress, the National Diet Library of Japan, royal libraries such as the Royal Library of Denmark, and Sweden, special national libraries such as the National Library of Medicine and National Library of Agriculture in the US These libraries constitute part of a nation’s sovereignty and pride (Hoare, 2007, pp.155-156).

Most countries have national or state library, which is maintained by government grant and by fee-based services offered by the library. National
libraries usually receive by legal deposit one free copy of each book or periodical published in the home country. Legal deposit libraries, which are basically archives of copyrighted works, have existed in some form or another since the sixteenth century. Countries such as France and Britain have national libraries with ancient roots in royal libraries; most well established national libraries exist in Russia (1810-11), Brazil and Argentina (both 1910), Belgium (1837) and Germany (1870). Also apart from preserving the nation’s literature, national libraries have generally collected material from other countries. Notable collections include the British Library’s India Office Library and Records and the extensive collection of Russian, Chinese and Japanese books held by the Library of Congress. The importance attached to collecting comprehensive ranges of foreign publications can be seen in the statistics that two-third of publications held by the US Library of Congress are in languages other than English (Feather & Sturges, 1997, op. cit.).

National libraries throughout the world have taken on the important mission of managing their national bibliography, basically viewed as an official record of a nation’s intellectual heritage or of its publishing products. Production of national bibliographies, however, is closely bound by legal deposit, a system that varies significantly between countries worldwide as discussed in Chapter (4), section (2.6). The most important obstacles facing national bibliographic services is the increase in digital publication as, traditionally, digital resources have neither been covered by legal deposit legislation nor subsumed in national bibliographies. As a result, the role of national libraries in producing the national bibliography will become increasingly dependent upon the generation and maintenance of links.
between publishers and other agents in the bibliographic information chain (Day et al. 1999).

IFLA Report (1992, op. cit.) provides that libraries over the world have to deal with the fast growing numbers of digital materials that need to be safeguarded. Publications in digital form, online or CD, digitised images, and born-digital objects need to be preserved and kept accessible. Safeguarding digital heritage is a major issue, especially for national libraries, because of their legal task of preserving the national heritage of a country.

Based on the above IFLA Report statement, the Sudan National Library (SNL) should adopt a new and urgent policy to establish the Sudan Web Archive project.

Reference to Article 90(1) of the 1998 Sudan Constitution, the President of the Republic issued an Interim Decree of the 1999 National Library Act and approved by the National Assembly. Article (2) of the Act defines the term ‘document’ as follows:

'any book or written paper or report or file or photo or sound or visual recordings or drawing or map or musical note or any material related to library collection and it should be part of governmental work or having national value brought to the library and registered as its collection'.

This definition corresponds to some degree to the NRO definition though there is no mention of any digital material including content published on the Web. The legal deposit of the materials defined above relies on the 1996 Copyright Act deposit, which has no legal mechanism to oblige publishers and individual composers to deposit their
intellectual products in the SNL in order to assist in achieving its objectives that are stated as follows:

- collects, preserves, and identifies the print national intellectual heritage;
- publishes the current national bibliography and maintains the union catalogue;
- administers the ISBN\textsuperscript{70} system;
- publishes the bibliography of the related documents produced abroad;
- manages public and school libraries;
- develops information resources in the Sudan;
- conducts studies and researches with aims to develop the library science;
- acts as the ISBN agent;
- maintains union catalogues for all state and federal public libraries;
- establishes a central unit for serials of the Sudan;
- provides the selective dissemination of information and current awareness services;
- collects and preserves documents, books, manuscripts, serial publications, microfilms, recordings, movies, and others, which is related to national or Islamic or Arabic or African or humanitarian heritage;
- provides library and information services for students and researchers\textsuperscript{71}.

The departments of the Library composed of the bibliographic control (the national cataloguing unit), department of public school libraries, and the information resources development\textsuperscript{72}.

\textsuperscript{70} International Standard Book Number

\textsuperscript{71} نفس المصدر السابق ونفس الصفحة.

\textsuperscript{72} المكتبة الوطنية، 2008. مشروع المكتبة الوطنية السودانية. اختراع: وزارة الثقافة والإعلام.
It also establishes strategic plan for the years 2007–2011 with the aim to collect, preserve and provide the national heritage in its all forms to all community users\(^73\).

It is managed by the National Library Board, which is formed by a decision of the Council of Ministers upon recommendation of the concerned Minister from a National Librarian and a number of qualified and experienced members including representatives of public libraries and the National Record Office. The Board’s duration is two years and reports to the Minister\(^74\). However, the responsibility of building the Sudan Web Archive is considered to be an important role that the Library should take the lead of. Any preservation programmes in the country are linked directly or indirectly with the national legislation as will be investigated in the below part.

### 4.4 The Intellectual Property Administration

It is apparent that IPRs have become an essential element in economic and cultural policy in a world in which the source of wealth is increasingly intellectual, as opposed to physical capital in which markets are distributed across the globe. For the purposes of this study, a consortium of the coming relevant institutions is recommended in order to co-operate with the Sudan National Library for preservation of Web-based content.

In the IPRs context, the Republic of the Sudan is a party in regional and international agreements on IPRs as follows:

- The Convention on Establishment of the WIPO since 15\(^{th}\) February 1974;

---

\(^73\) المكتبة الوطنية، 2008. مراكز خطة المكتبة الوطنية للعام 2008م. ازدحم:

\(^74\) السودان. قانون المكتبة الوطنية ( نفس المصدر السابق ولكن الصفحة).
• The Lusaka Agreement on Establishment of the ARIPO since 9\textsuperscript{th} December 1976;
• The Paris Convention for the Protection of Industrial Property since 16\textsuperscript{th} May 1984;
• The Madrid Agreement on International Registration of Trademarks since 16\textsuperscript{th} April 1984;
• The Patent Co-operation Treaty since 16\textsuperscript{th} April 1984;

There are several institutions that administer IPRs in the country. It is vital to discuss their roles and how they could assist in harmonising their national legislation for Web legal deposit purposes.

4.4.1 FCLAW

The Federal Council for Literary and Artistic Works (FCLAW) has been established in 1991 in accordance with the Decree No. (447) of the Council of Ministers at the Ministry of Culture, Youth and Sports. Its objectives are to administer the 1996 Copyright and Neighbouring Rights Act No. (54) which repeals the 1974 Act\textsuperscript{75}. The Council also administers the 2000 Literary and Artistic Works Act that abolishes the 1991 and 1993 Acts\textsuperscript{76}. To the moment, these laws do not incorporate provisions authorising preservation of Web-based content.
4.4.2 RGIP

The Registrar General of Intellectual Property (RGIP) agency has been established as part of the Ministry of Justice. Its objectives are to administer the industrial property rights. It also administers the 1931 Trademarks Ordinance as replaced by the 1969 Trademarks Act, the 1971 Patents Act (The Sudan, 1969, op. cit.) and the 1974 Industrial Design Act (Sudan Industrial Design Act, 1974).

4.4.3 CWTOA

As part of its mission to accede to the WTO, the Commission on the World Trade Affairs (CWTOA) administers the TRIPS agreement. The agreement obliges member states to legislate national laws that ensure what is described as minimal protection of IPRs, (Madar Research Group, 2006). The Government of the Sudan has applied to join the WTO in 1994 and submitted its memorandum of foreign trade regime in 1998. Since that date, it has received many questions of which some have been answered and some are await to be answered (Basbar, 2001).

To accede to the WTO in order to gain benefits of the global economy, the Republic of the Sudan has submitted a Legislative Plan with a request that it could be circulated to members of the Working Party. This is a time frame Tentative Legal Action Plan (see Table 4.5) intended to align the Sudan’s legislation with the WTO’s laws and disciplines (Working Party on the Accession of Sudan, 2004). Such alignment, according to the researcher, need to consider exemptions of the library and information services in their attempts to preserve the national digital heritage because it will be difficult to wait to amend these laws if these exemptions are excluded and it will be too late to
preserve the fading Web-based content. Also librarians and information professionals need to be involved while formation or amendments of laws concerning provision and access to information are made.

4.4.4 PPPMNC

A review of the 2004 Act of the Press and Printed Press Material National Council (PPPMNC) reveals that it does not include any provisions for preservation of the daily newspapers published on the Web. Article 25(e) of the Act obliges the press institutions to establish information centres but without stating clear objectives and functions including preservation of press and printed press material on the Web77. The above-mentioned agencies need to review and update their legislations and permit copying for preservation of Web-based content.

77 Sudan. قانون الصحافة والطبوعات الصحفية لسنة 2004م. أخر طوم: المجلس القومي للصحافة والطبوعات.
Table (4.1) Sudan Agenda for Enacting Laws for WTO Conformity.

<table>
<thead>
<tr>
<th>Law</th>
<th>Agreement</th>
<th>Status of Draft</th>
<th>Expected period for Enacting &amp; Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copyright &amp; Neighbouring Rights Act</td>
<td>TRIPS</td>
<td>Draft Completed by the Ministry of Justice</td>
<td>2 years</td>
</tr>
<tr>
<td>Trademarks Act</td>
<td>TRIPS</td>
<td>Draft is prepared by the Ministry of Justice</td>
<td>2 years</td>
</tr>
<tr>
<td>Patents Act</td>
<td>TRIPS</td>
<td>Draft is prepared by the Ministry of Justice</td>
<td>2 years</td>
</tr>
<tr>
<td>Industrial Design Act</td>
<td>TRIPS</td>
<td>Draft is prepared by the Ministry of Justice</td>
<td>2 years</td>
</tr>
<tr>
<td>Protection of Undisclosed Information Act</td>
<td>TRIPS</td>
<td>Draft is prepared by the Ministry of Justice</td>
<td>4 years</td>
</tr>
<tr>
<td>Business Name Act 1931</td>
<td>TRIPS</td>
<td>Draft is prepared by the Ministry of Justice</td>
<td>[N/A]</td>
</tr>
</tbody>
</table>
4.4.5 TIPC

The Trade and Intellectual Property Court (TIPC) has been established in 2002 in order to administer the civil cases regarding companies and trademarks, business names, the intellectual property claims and other cases.

4.5 Proposal for Web Legal Deposit

As reviewed in the above part and as defined in Chapter (1), section (1.8.2), the first step in legalising copying for preservation purposes is to propose establishment of a Web legal deposit law. The literature review results (see Chapter (1), section 1.9) show that studies have tended to call for updating the current legal deposit, copyright and other legal instruments and adapt them to the Web in order to enable national libraries to set up Web archives and preserve access to their domain space. These views are manifested in studies in the second dependent variable as of Charlesworth (2003, op. cit.) and Jacobsen (2007, op. cit.). In contrast, the result of study in the first dependent variable as of Smith (2004, op. cit.) discloses that reliance on legal deposit as enforced on print static material is no longer feasible on the Web. In Chapter (2), section (2.6.9) Yokoyama (2006, op. cit.) warns that the legal deposit system is inappropriate in the light of the fundamental principles of the legal deposit system (Yokoyama, 2006, op cit.). According to Knutsen (2006, op. cit.) the NDL has prepared legislation for the acquisition of Web resources. As investigated above and reference to the results in Table (5.3) in appendix (A), which provide 13% of the institutions rely on legal or voluntary deposit and 6% rely on copyright deposit, while 80% of institutions investigated rely on different other mechanisms, hence it is proposed in this
study to establish a separate Web legal deposit, why? Because it will legalise coping with challenges of frequent copying for migration and emulation strategies due to the dynamic hyperlinks, software and hardware dependency as opposed to the print materials (in this sense, preservation is meant to maintain long-term access). The proposed law should first define and identify the national domain boundaries and their relevant links. This is because the concept of the digital ‘document’ differs fundamentally from its print counterpart as illustrated in Chapter (1), section (1.9.1) and Chapter (2), section (2.3.9).

4.6 The Preservation Strategies

There is no worldwide acceptable definition for the word ‘strategy’. Its origin goes back to the Greeks who were conquered by Alexander the Great and his father. In fact, the word ‘strategos’ was used to refer to specific rank of those of the commander in chief of armies. Later it meant ‘the general’s skills’, or the ‘psychological aptitude and character skills’. About 450 BC it referred to administrative skills (management, leadership, oratory and power) while at the time of the Alexander (ca. 330 BC) the word was used to refer to the ability to apply force, overcome the enemy, and create a global, unified system of government. In the realm of management, ‘strategy’ means a pattern or plan that encapsulates the main objectives and policies of an organisation while also establishing a coherent sequence of actions to be put into practice (Cunill, 2006, pp.1-2). The strategy often falls into one of four categories:

- ‘Plan’ which focuses on the ‘how’ approach and aims to get from here to there and links between ends and means.
• ‘Pattern’ which focuses on actions over time to reach ‘higher ends’;
• ‘Position’ which focuses on reflection of decisions to offer particular services in a particular community.
• ‘Perspective’ which focuses on vision and mission, a view of what the institution should become (Nickols, 2000).

The rationale for preservation of Web-based content ranges from national library to another but the essence remains the same. Article (6) of the revised UNESCO draft Charter (2006) on the ‘Preservation of the Digital Heritage’ asserts that strategies and policies for preservation of digital national heritage should be developed taking into consideration certain factors such as urgency, local circumstances, unavailable means and prospective projections. Crucial to strategy is to set up standards, compatibilities, and resource sharing in co-operation between authors, right holders and the relevant institutions.

The preservation strategy also described (Verheul, 2006, op. cit., p.20) as the procedures that enable permanent access to stored digital material. It refers to all techniques that provide more than would be obtained by merely preserving the digital objects and never look at them again.

Technological change is a permanent trend of the hardware industry, just as it is in multimedia and software publishing (Miura, 2007). Also digital information is fragile; it can be corrupted or altered, intentionally or unintentionally. It may become unusable as storage media and hardware and software technologies change. Format migration and perhaps emulation of current hardware and software behaviour in future hardware and software platforms are strategies for overcoming these challenges (NISO, 2004, op. cit., p.2).
Possible solutions for preserving digital content include:

- ensuring that a copy of every preserved document survives as long as it might interest potential readers;
- ensuring that authorised consumers can find and use any preserved document as its producers intended, without difficulty from errors introduced by third parties that include archivists, editors, and programmers;
- ensuring that any consumer has accessible evidence to decide whether information received is sufficiently trustworthy for his applications;
- hidden information technology complexity from end users (producers, curators, and consumers);
- minimising the costs of human labour by automatic procedures whenever doing so is feasible;
- enabling scaling for the information collection sizes and user traffic expected, including empowering editors to package information so as to avoid overloading professional cataloguers; and
- allowing each institutional and individual participant as much autonomy as possible for handling preserved information, balancing this objective with that of information sharing (Gladney, 2007, op. cit., p.12)

The Sudan National Library needs a well-defined strategy to determine the value of the services it provides to its patrons, so it can assess their satisfaction accordingly. This strategy should be aligned with policies of the e-government and e-commerce. The e-Government is looked at as a process that the citizenry, in pursuit of its governance, conducts over a computer-mediated
network (Scholl, 2006), while the e-commerce is looked at as technology-mediated exchanges between parties (individuals or organisations) as well as the electronically based intra- or inter-organisational activities that facilitate the exchanges (Rayport, 2002, p.1).

These ‘technology-mediated exchanges’ have enabled some companies and countries around the world to get use of the knowledge economy. For instance, in May 1997 the Amazon, which is founded in 1995, engaged in a fierce competitive battle with the Web-based subsidiary of Barnes and Noble, the largest bookseller in the US, over dominion of the online bookselling market. While Wall Street analysts did not expected Amazon.com to record a profit until 1999, Amazon’s market valuation at year end 1995 was approximately USD$2.00 billion, with its stock price at USD$7.5 and USD$10.00 billion book sales (Chevalier & Goolsbee, 2003, p.4).

When Marshall McLuhan suggested that electronic interdependence was recreating the world in the image of a global village, he was writing in the early 1960s, long before the rise of the Internet revolution. Indeed, in this first decade of the twenty-first century, the world has become much smaller and communities more closely knit in a way that a person from Cape Town can pay and receive a book from Oxford within minutes via downloading. Investment in ICTs by firms in sectors such as electronics manufacturing, media, wholesale, retail trade, finance and insurance, and business services contributed over 1% to Australia’s productivity growth from 1996-2001, as much as 1.3% in the US and 0.73% in Ireland (OECD, 2004).

Professionals working on preservation generally agree on the problems or challenges they face. They are typically described under three categories: hardware obsolescence, software dependence, and storage medium deterioration. While all three are
eventually affect the long-term survival of digital objects, most experts agree that it is software dependence, or 'the fact that digital documents are in general dependent on application software to make them accessible and meaningful', that presents the greatest challenge. There are major strategies need to be maintained in order to overcome these problems discussed as follows:

### 4.6.1 Selection

Though it is difficult, not to say hazardous, to decide on the contemporary or historical research value of an item, the question should at least be raised. As an example, should all annual Web-based reports of health care mainly containing medical records, be deposited and preserved? Should all publications of all ministries, mainly containing educational, statistical and administrative information, be preserved?

There are no standards to be followed or even recommended, since the answer of the above raised questions basically depends on the Sudan National Library’s policy and its legal deposit jurisdiction.

In fact, the Web consists of huge amount of digital garbage but in the same time it consists of tremendous volume of valuable content useful for future research and development. According to Guedon (2003), the policy of peer-reviewed research for digital scientific content has become a reality on most commercial databases available on the Web.

A report by the OECD (2004, op. cit.) organisation shows that e-mail is by far the leading use of the Internet service, followed by information search, with purchase of goods and services and other online transactions and other advanced forms of management, lagging well
behind. In 2002, study (OCLC, 2004, op. cit.) estimated that 31 billion e-mails travelled the Internet daily and in 2006 60 billion e-mails has been sent over the Internet.

Gladney (2007, op. cit.), however, mentions several threats to preservation of digital content as media, hardware and software failures, communication channel errors, network service failures, component obsolescence, operator errors, natural disasters, external attacks, internal attacks, economic and organisation failure.

Reference to a paper prepared by Frieder and Zittrain (2007), unsolicited e-mail or spam is account for 80% of all Internet e-mail traffic with estimates topping at 1.6 billion messages per week. In 2004, estimates put e-mail spam as high as 40 or 50% of the Internet traffic but if spam is excluded, about 22 billion e-mail messages delivered to e-mail accounts that have content meaningful to the recipients (OCLC, 2004, op. cit.).

Precautions should be dealt with when selecting Web-based content for long-term preservation. Web-based content is particularly vulnerable to external attacks. In 2001, for instance, the Nimda worm, which took down 150,000 computers, and the Code Red virus, which struck more than 12,000 Websites in the US, hit the Web hard. In June in the same year, Microsoft had issued a patch to protect its attack-vulnerable Internet Information Server (IIS) software, which is used by approximately 16 million Websites (Kenney et al. 2002, op. cit.).

Without doubt, these threats affect the process of the selection and collection development of Web-based content. Though the survey results show that only 11 (3%) (see Table (5.5) and Figure (5.2) in appendix B)
respondents predict that the Internet may be hampered by disasters though the fact discussed in Chapter (2), section (2.2) states that this network is designed originally to recover from problems such as power failures, interruptions in communication lines in the case of a nuclear disaster.

Based on library principles, Web-based content can be analysed for selection in terms of the Ranganathan’s\(^ {78} \) five library laws considered as essential components of the insight behind the library functions:

- Books are for use.
- Every reader has his or her book.
- Every book has its reader.
- Save the time of the reader.
- The library is a growing organism.

These principles were developed in India in 1931 and accepted worldwide as the theoretical basis of ideal library services (Ranganathan, 1931). Inspired by such principles, Noruzi (2004, op.cit.), has introduced the five principles for implementing the selection policy on the Web environment as follows:

- Web resources are for use.
- Every user has his or her Web resource.
- Every Web resource has its user.
- Save the time of the user.
- The Web is a growing organism.

In order to derive these principles putting into their consideration that the Web consists of published content from any entity or person and that the quality of information or the value of

---

\(^ {78} \) Dr Shiyali Ramamrita Ranganathan, 1898-1986, has many contributions including philosophy of library classification published in 1973.
knowledge is opaque, due to the lack of any kind of peer review methods as discussed in the coming part.

Preservation is a process that aims to save the memory of knowledge for the coming generations of national library users and the society in general. Previously, long-term preservation of single items was extremely expensive and carried out on only the most important and valuable pieces. As a result, libraries turned to apply another options such as microfilming and digitising to save the intellectual content. Studies (Pilette, 2005, p.31) revealed that the estimated cost of microfilm for an average volume in 2003 is about USD$118.00. Once filmed, there is the cost of maintaining two negatives, the mater (master) and a printing negative, and a positive viewing copy as well as the machines to view the film. Digitisation (cf. Chapter (2), section (2.3.7) though generally less expensive, can still cost about USD$74.00 per average volume. The ISO organisation (1995) estimates that the cost to recreate a megabyte of the lost data is at USD$1,250. But preservation of Web-based content is less expensive because it relies on digitised content or content that already born-digital and published on the Web.

A survey of the 21 members of the Digital Library Federation revealed that 40% of their costs for digital libraries in 2000 went for commercial content. While the big-ticket content were scholarly e-journals that libraries licence rather than own as discussed in Chapter (2), section (2.4.5). But none of such libraries appear ready to forgo access to the licenced content because its long-term accessibility might be in question (Kenney et al. 2002).

This is because digital content creators or publishers constantly update their databases and
do not care much about retaining previous versions (Ershova & Hohlov, 2002, op. cit.). For the selection purposes of this study, it is not enough for content to be present within the system, they have to be readily accessible ('every reader has his or her book', in Ranganathan's terms), potentially interested readers have to be aware of them ('every book has its reader'), and the system for matching supply and demand has to be efficient ('save the time of the user'). For ensuring that Ranganathan's principles are eligible on the Web environment Lennart Bjorneborn has suggested in his thesis ‘five laws of web connectivity’ (Bjorneborn, 2004).

Due to the dramatic increase in expenditure on online content in academic libraries, selection of Web-based content should include coordination between key players such as the Sudanet, Zenanet), publishers and other national information institutions. The Sudan National Library is the most appropriate institution capable for the mission. It should build a Web depository in order to maintain a sustainable access to Web-based content.

4.6.1.1 Quality of Content

Quality of content published on the Web is often criticised for being unreliable and lack credibility to be utilised for decision-making, education and research and development. This is because there are no quality assurance or governance policies in the Web publishing. As a result, the proposed Sudan Web Archive should adopt a clear collection and capture policy in order to guarantee quality assurance of its output. The International Organisation for Standardisation, (ISO)-(8402) dealt with quality assurance as those planned and
organised activities required to provide confidence of a product or service.
But it is essential to ask a fundamental question why applying quality assurance policy on preservation of Web-based content? The possible answers could be because there is:

- An increasing demand for accountability internally and externally;
- A competitive e-education market;
- Developments in e-education, e-commerce and e-governance;
- A growing volume of accessible information;
- Reconciliation of the needs of students; undergraduates and postgraduates;
- Recognition of the limitations of technology for many students;
- Balancing the often-competing needs of a diverse user group.

During the past decade a small but growing number of national libraries have established Web archive programmes as discussed in Chapter (1), Section (1.9.1.1). These projects, as Phillips (2005) describes, have taken one or more of four main approaches for building their national collections:

- Selective preservation, for example, the Web archives of the National Libraries of Canada, Japan, and Australia;
- Periodic harvesting of the country’s entire Web domain, exemplified by the archives of the Nordic countries, including Sweden;
- Thematic collecting, exemplified by the Library of Congress’s MINERVA collections
of the Elections 2000 and of September 11th, 2001;
- Deposit collections, such as STORS79 at the State Library of Tasmania; and
- The e-Depot at the National Library of the Netherlands.

The International Council on Archives (ICA) (2005) identifies five general criteria for the selection purposes. These criteria are authenticity, completeness, accessibility and understandability, processability, and reusability. Selection of Web-based content requires an extensive Web navigation and administration skills. As a result, librarians need to utilise automated tools to locate reliable content and to track and analyse their usage patterns. Cooley et al. (1997), however, provides that such tools give rise to the necessity of creating server-side and client-side intelligent systems that can effectively mine for knowledge. He also looks at 'Web mining' as the discovery and evaluation of useful content on the Web. This includes implementation of automatic search of information resources available on the Web.

Acquisition and collection development in libraries are often done through traditional means such as purchases, donations, right transfer or legal deposit. According to James H. Billington (2002,p.2), the Library of Congress collected 29 million books and 103 other library materials (manuscripts, motion pictures, sound recordings, maps, prints, photographs) in two centuries. Nowadays, it takes only 15 minutes for the global community to produce the same amount of content in digital form.

79 Stable Tasmanian Open Repository Service.
Reference to the vision of the ADCIA of the role of library content in Chapter (2), section (2.3.13), the issue of quality is essential in Web archiving for it support activities of the national content industry.

4.6.2 The Harvesting

Unlike the traditional library acquisition, Web harvesting is the new term used for collecting Web-based content. Henceforth, harvesting such kind of content requires an obvious plan for preservation as an integral part of the National Library’s mission, and preservation planning should be part of its overall strategic plan. However, preservation planning could be looked at as:

- A process by which the general and specific needs for the care of collections are determined, priorities are established, and resources for implementation are identified (starting from here to there).
- It is main purpose is to define a course of action that will allow the national library to set its present and future preservation agenda (a pattern).
- It identifies the actions the National Library needs to take care of and those it probably never takes care of so that resources can be allocated appropriately (IFLA-CDNL, 2006, p.12).

It is important to mention that in the EU, there are 30 members of state national libraries have legal deposit law requiring deposit of some or all types of digital publications including Web-based content. About 13 libraries are testing Web preservation project at the time inquired
(see the questionnaire analysis results in appendix A). More than eleven national libraries preserve Web-based content of the entire national domain name space, for instance, ‘.fk’, ‘.uk’, nine of these libraries collect resources published outside their national domain space, i.e., ‘.com’, ‘.org’, ‘.net’. Material collected from these sites ranging from public domain: seven libraries and through purchases: two libraries (Stoklasova et al. 2005).

4.6.2.1 Harvesting the Sudan Domain

The ccTLD of the Sudan ‘.sd’ is chosen as the setting for harvesting the Web-based content. The second level domain name available under ‘.sd’ includes many other domains as shown in Table No. (4.2).

Table No. (4.2) the registered DNs categories on the ‘.sd’ Sudan ccTLD domain as administered by the SIS.

<table>
<thead>
<tr>
<th>.sd</th>
<th>Companies, Organisations or Individuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Com.sd</td>
<td>Companies</td>
</tr>
<tr>
<td>Net.sd</td>
<td>Network Providers, ISP</td>
</tr>
<tr>
<td>Org.sd</td>
<td>Sudanese NGOs</td>
</tr>
<tr>
<td>Edu.sd</td>
<td>Sudanese Universities and Colleges</td>
</tr>
<tr>
<td>Med.sd</td>
<td>Medical</td>
</tr>
<tr>
<td>tv.sd</td>
<td>Media</td>
</tr>
<tr>
<td>Gov.sd</td>
<td>Sudanese Government and Ministries</td>
</tr>
<tr>
<td>info.sd</td>
<td>Newspapers, Information &amp; Media</td>
</tr>
</tbody>
</table>

Within the Web legal deposit jurisdiction, the Web harvesting in the BnF, for instance, applies three levels as follows:

- Putting out of scope all DNs that are not obviously containing resources belong to the French culture;
- Focusing on the specific French national domain names (.fr, .pf, .wf, .pm, .re, .tf, .ad, .yt);
- Harvesting all the rest (Lupovicic, 2005, op. cit., p.2).

To build the Sudan Web Archive, this study advocates similar procedures as of those of the BnF. Reference to the Sudan National Information Centre (NIC) the total number of the Sudanese Websites is 722 (NIC Website). The researcher noticed that most of these sites contain static HTTM-based content and it is hardly to find interactive multimedia items. The NIC provided no details on the numbers of the national Web space. As shown in Table (4.2) above, the scope of building the Sudan Web Archive’s collections is to focus on the entire content published on ‘.sd’ While other foreign domains that publish relevant content about the country might constitute a big challenge to maintain full coverage of the Archive holdings due to IPRs barriers as discussed Chapter (2), section (2.4).

The basic method in the harvesting process is an automatic crawling (crawlers are discussed later in section (4.7.2.1). The entire surface Web and of the deep Web can be harvested in such a way, depending on the technical problem the crawler faces and how it has been tuned. But generally speaking, an automatic crawler mainly collects the surface Web, creating a map of the Web.
A second and complementary harvesting method is to focus the automatic crawling on a particular set of Websites, for instance, at ‘.sd’ domain, and ask the crawler for a focused harvesting in depth.

These two last methods are top down oriented. In order to get the whole Web picture, the Sudan Web Archive should combine the periodic automatic harvesting, with a bottom up oriented technique, which assists to complement the collections with the deposited content by publishers and to recreate the links with the surface Web already harvested. Those deposits need to be processed partly manually (via e-mail attachment or providing link to huge volume of content as in PANDORA), or partly automatically (as in BnF Web Archive). The deposit can present advantages in terms of preservation insurance, but it is time consuming process, starts by the analysis of the Website, which may requires five persons per day on average (Philips, 2004, op. cit.).

Nevertheless, it is important for the Sudan Web Archive to collect and preserve the content on the Web that is appropriate for the national cultural memory, and this presents challenges on a scale as large as the Web itself. There are formidable technological challenges, common to all digital documents, of course. But beyond those, the problems start with capture of the Web. Web harvesting, an approach used to create a ‘snapshot record’ of the Web, can capture static HTML pages, the deep Web, where much of the complex and culturally rich materials reside, is often inaccessible to harvesting technologies.

Even the surface Web is closed to harvesters in many cases because the materials require a licence or other authorisation to login. Some
studies (IIPP, 2002, op. cit.), however estimate the average Web page contains 15 links.

But how does one define the boundaries of a Website? This is among many questions that emerge from early experiments intended to capture content from the Web as discussed in Chapter (2), section (1.9.3). National libraries, for example, have lasting experience in the selection of content that has long-term cultural value among an abundance of compelling print material. In this sense, the Web is familiar to librarians as a medium that contains text, graphics, and images, and that indifferently carries content as diverse as Shakespeare anthologies and screenplays, photographs, music, manuscript, and maps (Bailey & Thomas, 2006, op. cit.).

Libraries have built collections of great merit and they need to make difficult choices of selection among such materials on paper. The challenge of selection from the Web may turn out to have similar conceptual complexities, but the scope of content is vaster. Most experts agree that identifying and capturing content of enduring value on the Web is the most formidable challenge for preservation it. It may be advisable to start with ‘published’ content, such as online books, theses, dissertations, journals and other items that have scientific value in the analogue realm for acquisition and preservation, and negotiate the deposit of the proprietary content. Another area of promise is to capture government information that is on the Web where the online versions have superseded print versions as official Websites are going digitally and establishing e-governments services (Day, 2003, op. cit.). It is also important for the Archive to capture
foreign Websites having content by the Sudanese or about the country in general.

The above two strategies of selection and harvesting are concerned with ensuring quality while the next strategies are essential to extend the life of the selected and harvested content for sustainable access. They mainly prevent obsolescence of software and hardware and other technical problems as stated in Chapter (1), section (1.1.1).

4.6.3 Technology Preservation

A process made to preserve the original software (and possible hardware) that has been used to create and access the content. It also involves preserving both the original operating system and hardware on which content is operated (Feeney, 1999).

4.6.4 Emulation

This is a process made to overcome hardware and software obsolescence, whilst retaining aspects of ‘look and feel’ and functionality. The emulation process involves encapsulation of content, original software, specifications and reverse engineering of original software, in order to develop emulator specifications or software (Granger, 2000).
4.6.5 Refreshing

The purpose of this strategy is to overcome storage media deterioration, obsolescence and involves periodic copying of bit streams from one physical medium to another (Ayre & Muir, 2004).

4.6.6 Migration and Conversion

A process made to overcome hardware and software obsolescence, without necessarily retaining ‘look and feel’ and functionality. This is done through conversion of content from one generation of computer technology to a subsequent one. It involves content format conversion, recording and saving information about the original software environment (Mellor et al. 2002).

Bescos (2004, p.3) suggests that migration is broader and richer mechanism than refreshing. It is a set of arranged tasks designed to conduct a periodic transfer of digital materials from one hardware or software configuration to another or from computer generation to a subsequent one. It aims to preserve the integrity of the digital objects and to retain the ability for clients to retrieve, display, and use of these objects in the rapid change technology market.

While the conversion strategy is viewed as the automatic processes of transfer of electronic documents from one application environment to another with little or no loss in structure, content or context even though the underlying bit stream is changed (Dollar, 1998). A typical example of converting electronic document is moving them from a software environment or application to another, such as converting a file from WordPerfect to Microsoft Word. Dollar limits the migration of electronic documents to
narrow circumstances in which neither backward compatibility nor export or import gateways exist between the legacy systems that contain the documents and the new application platform. In Dollar's view, the primary difference between migration and other digital preservation strategies is that migration involves proprietary legacy systems that lack export software functionality and the only way known to migrate the documents along with essential software functionality to an open system is to write special purpose code or programmes. Migration strategy is critised as the most complex and costly of the digital preservation strategies.

Because it is a complex preservation strategy, it presents serious challenges to the content when attempting to preserve its authenticity. Of particular note is the potential loss of structure and functionality resulting in the inability to truly represent, use and interpret the content. As a result of these issues, opponents of this strategy argue that it is essentially an approach based on wishful thinking (Kent et al. 2003, op. cit.).

4.6.7 Long-term preservation

Digital preservation consists of processes aimed at ensuring the continuity and accessibility of digital materials. National libraries worldwide have formed preservation strategies for materials that intended to be transmitted to subsequent generations of information seekers, researchers and scholars. For paper-based documents, provision of adequate storage conditions was the best means to help ensure that materials would remain readable far into the future. With the advent of digital technology most publishers do their works on databases. Some of these works may be printed on
paper, but much of it, particularly databases, geographic information, scientific data sets, and Websites, maps and atlases exist only in electronic form. At the same time, traditional forms of publications have changed significantly and, as a result, create new challenges. For instance, publishers of e-journals licence their content to libraries, but libraries do not own that content and they may not have legal rights to capture digital content to preserve it as discussed in Chapter (2) section (2.4).

To ensure longevity of Web-based content, Webb (2003) argues that in order to achieve long-term preservation of digital materials, these materials need to be viewed as physical phenomena, logical encoding, conceptual meaningful material and as sets of necessary digital objects accessible to future generations.

To answer the question of how long national libraries need to preserve Web-based content? Estimations (ICA, 2005, op. cit.) provide that some e-documents could be preserved for permanent period of time, or fixed period of time, which greatly exceed the lifetime of the software and hardware such as personal records, which might be retained for 75 or 100 years from their creation, or such other periods as law may provide. It also includes preservation for indefinite, but not infinite, time periods which can be expected to exceed five years, such as building designs and drawings which will be retained for at least as long as the life of the building to which they refer to.

More and more valuable content is born-digital and can be managed, preserved, and used in digital form. In the last decade, researchers (Kanehisa & Bork, 2003, p.305) have mapped significant portions of the human genome. Advances in biomedical research depend on
building and preserving complex genomic databases. Research in biodiversity and ecosystems, global climate change, meteorology, and space science – to name only a few fields – is built on the ability to combine vast quantities of digital information with complex models and analytical tools.

The entertainment content industry has shifted rapidly to digital masters for recorded music, movies, games and interactive television as discussed in Chapter (1), section (2.3.13). These provide critical resources for research, historical documentaries, and cultural coherence. Audio, film, and video recordings are replayed, rebroadcast, and reviewed as a source for entertainment and vital connections to the past (Gafar, 2006a).

Information systems used in business and government generate enormous quantities of digital information, some of which is worth saving for a long period of time. The aeroplane industry, for instance, depends on software systems to design, manufacture, and maintain complex commercial products. For safety’s sake, design specifications, records of manufacturing processes, parts inventories, maintenance records, and performance data, much of which is in digital form, should be kept as long as a particular model of aeroplane is in service – a period that can exceed fifty years. The food and drug administration, for example, requires pharmaceutical companies to file new drug applications electronically along with documentation of research protocols, tests, and clinical trials. These digital documents should be kept at least as long as a drug is available. Medical records that may be needed for an individual’s entire lifetime are becoming electronic. Citizens’ rights, such as eligibility for social security benefits, are
documented in databases that accumulate data through each individual’s working life (NSF & LC, 2003, passim). E-government and e-commerce could flounder if better methods are not found to identify and preserve digital records that are essential for keeping the business running and for maintaining accountability. Stringent requirements for authenticity and integrity permeate this problem (Gafar, 2006b).

The above-mentioned strategies are considered as a ‘pattern’, which focuses on preservation actions over time in order to reach higher ends, i.e. ‘access to information’. To reinforce these ends, the below section will provide a ‘plan’ in order to focus on the ‘how’ approach for getting from here to there.

4.7 Web Archive Operations

Further to the definition of the ‘Web archive’ in Chapter (1), section (1.8), there are five phases in the development of information retrieval system usually described as design, development, implementation, operation, and maintenance (Koneru, 2005, p.6).

The design phase is the stage in which plans, models and workflows are sketched out and documented. The development phase is the stage in which the digital repository is launched. The implementation phase is the stage in which the system is implemented in the National Library, and linked to other systems and processes in the organisation. While the operation phase is the stage in which the digital repository is put into practice for handling the digital objects, both current and in the future (Ibid).

As discussed in Chapter (1), section (1.9.1.1), it is important to notice that the existing Web initiatives have drafted international legislation
to regulate the cyberspace. Such legislation also intends to govern mass proliferation of Spam, counterfeit products, piracy, pornography, terrorism and cybercrimes. These efforts intend to curb the so-called immoral 'World Wild West' from penetrating into conservative societies. This trend has been presented at the WSIS Summit in Tunisia though vetoed by the US and its allies who are in favour of keeping such tool open to all may be for protecting their economic interests. Many users in the Internet community favour this arrangement, viewing cyberspace as a virtual world where anarchy reigns over supreme. But as in any society, while anarchy may benefit a few in the short run, eventually order should be restored for the greater good of all. Therefore, details of the system description need to be clearly designed to administer and utilise benefits of the national Web space (Albert et al. 1999, op. cit).

In Africa less than 12% of the population in South Africa have access to the Internet, far less have access to commercial digital archives. The figures are much lower in the rest of the continent (Nicholson, 2007, op. cit.). This is because commercial archives effectively serve an elite educated audience, mostly in industrialised countries such as the Sudan Archive at Durham University, the Sudan collections in the University of Bergen and the School of African and Oriental Studies at the University of London.

The main objectives of the legal deposit law, as spelled out in the UNESCO Guidelines for legal deposit legislation (cf. section 2.4.8.1), are to enlarge library collections, build statistics, and enable book exchange (UNESCO, 2000, op. cit.). In addition to the UNESCO legal deposit objectives mentioned above, the legal deposit also assists in preparing the national bibliography in conventional print form or the national metadata schema on the Web.
Article (9) of the UNESCO Charter (2003, op. cit.) on the ‘Preservation of the Digital Heritage’ provides that digital heritage is culture-specific and unbound by time, location or format. It is essential to preserve and make it accessible in order to assure representation of all nations, cultures and languages.

Building a Web archive defers from the creation of the bibliographic or full-text database. This is because such environment is based on the ccTLD domain of certain country and its associated affiliates or beyond. The domain names are part of the preserved national Web-based content but are subject to IPRs as discussed in Chapter (2), section (2.4). The flowchart of the proposed Sudan Web Archive is shown in Figure No. (4.2) and its operations as in figure No. (4.3).

The major problem in building the Web archive is that on the Web environment the archive cannot preserve the content of information without preserving the associated software which enables hyperlinks to further content. According to Cerf (2005, op. cit., p.11), content is actually encapsulated in the middle of software.

Though such problem, the WSIS Action Plan agenda has advised that countries should develop a framework for the secure storage of archival documents and other electronic records of information (WSIS, 2003).

However, the major goal of preservation of Web-based content is to make information and knowledge available and accessible for the generations to come since their information requests and needs are different and depends on the situation, time and the type of research to be pursued. Socrates, for instance, believed that there is an important correlation between ‘virtue’ and ‘knowledge’, and that knowledge is necessary and sufficient for virtuous conduct. Two arguments used to support his idea are as follows:
All rational desires are focused on what is good; therefore if one knows what is good, he or she will not act contrary.

If one has non-rational desires, but knowledge is sufficient to overcome them, so if one is knowledgeable of goodness, he will not act irrationally (Reshotko, 2006).

Socrates held that to exercise the virtues properly, one should have knowledge of them. He felt that one can misuse a virtue; for example, one can have either too much or too little generosity, and neither one is virtuous. One should know how to use the virtue of generosity, as well as courage, honesty, and loyalty, (This idea is known as Socratic Intellectualism which was later revived after Socrates and practiced by the Stoics) (Ibid).

The researcher’s comments to the above arguments is based on the idea that knowledge extracted from preserved Web-based content could enhance and contribute to human virtue because such knowledge will be available and accessible to thinkers and scholars through the Web archive.

However, before setting up a plan for the proposed Web archive, it is better to look at the content lifecycle in order to draw final guidelines of the Web archive:

- Selection: (assessment and filtering plan)
- Harvesting: (applying harvesting tools)
- Analysis: (classification and cataloguing)
- Storage: (distributed backup and security measures)
- Retrieval: (creation of national bibliography and metadata)
- Disposal: (assessment and weeding plan)

The proposed Sudan Web Archive (SUWA) is to be built on the ‘.sd’ domain linked directly with the
metadata in addition to other four links. The metadata, in turn, is linked with the online public access catalogue (OPAC) in order to provide multiple access options to the Archive. The operations of the proposed Web Archive with their flowchart are provided in Figure No. (4.2).

Figure No. (4.1) Flowchart of the Proposed Sudan Web Archive.
Reference to Brugger (2005), the Internet is a very dynamic object. The fact that the sender or publisher renews the content on the Web or each copy of a Website at certain frequencies or intervals is a matter of space and of time. In order to preserve these updates it is wiser to know where and when something is updated.

Another aspect of the dynamics of the sender is the proliferation of senders (or publishers). As it is common for other media senders come and go, but what is different by the coming and going of senders on the Web is the ‘number’, the ‘nature’ and the ‘speed’ (Snelson, 2005).
Figure No. (4.2) Web Archiving Operations
4.7.1 Crawlers and Harvesting Tools

In contrast to acquisition and collection development in traditional libraries, harvesters or crawlers are software designed especially for collecting library digital materials published on the Web. A crawler is described as automatic equipment designed to collect Web documents. If a number of URLs identified as testbed, the crawler scans the documents, check the associated hypertext links then continues into a second harvesting stage; this operation continues until all qualifying documents are entirely scanned for retrieval (Hakala, 2000). The crawler is also described as an application that scans and archives Web-based content in compliance with a set of user-defined parameters (Hakala, 2004).

The harvester works in a similar way to a browser for it makes requests of a host files. Then it follows links within a site and collects files it finds and it is capable of harvesting database driven Websites, the deep Web content, such as library catalogues. But it cannot collect any content that is protected behind a password nor can it break or crack passwords (UK Web Archiving Consortium Website, op. cit.). The standard representation in the HTML or XML has led to the creation of the crawlers. In the case of Google crawler, it goes through some five billion pages, which constitutes only a fraction of all the content on the Web (Cerf, 2005, op. cit., p.14). The crawling process utilises the information extraction and indexing techniques for information discovery as applied in natural language engineering (Srihari et al., 2006).

The harvesting tools used in preservation of the Web vary in quality and function. The decision-makers need to select one the below crawlers after a thorough scrutiny.
4.7.1.1 Heritrix

A Web harvester used by the Internet Archive and considered one of the most important Web harvesters used for collecting Web-based content and distinguished by its use of the ‘.arc’ file format. This format allows for massive Web crawling efforts without reproducing the original site structure complexity (Seneca, 2006).

The name ‘Heritrix’ is derived from an archaic word for ‘heiress’, a woman who inherits. This is because it ensures access to the documentary heritage for the future generations. It is an open source, extensible Web-scale, archival-quality JAVA script software developed in the first half of 2003 and released in January 2004. It is a part of the Internet digital library project launched by the Internet Archive, which contains over 400 terabyte data. Its open nature aims at promoting collaboration, share experience and develop new features between the cultural institutions interested in preserving the Web (Mohr, 2004). Heritix is driven by XML configuration language, which supports complex crawl definitions and filtering. It also supports advanced customisation via JAVA plug-ins. The crawler includes a Web hosted control panel for managing and monitoring crawls (Marill et al., 2004). It operates at the Internet Archive, Czech Republic National Library and Slovenia National Library (Wiggins, 2005, op. cit.).

Setting up a crawler involves choosing and configuring a set of certain components to run successfully. Operation of the crawl repeats the following recursive process, common to most Web crawlers, with the specific components chosen:

80 Accessible at http://crawler.archive.org
• Choose a URI from among all those scheduled;
• Fetch that URI;
• Analyse or archive the results;
• Select discovered URIs of interest, and add to those scheduled;
• Note that the URI is done and repeat.

The Heritrix is composed of three most prominent components such as ‘Scope’, the ‘Frontier’, and the ‘Processor Chains’. Their functions are to define the crawler’s work. The function of the ‘Scope’ determines what URIs are ruled into or out of a certain crawl. It includes the ‘seed’ URIs used to start a crawl, in addition to other rules used to determine which URIs discovered to be scheduled for download (Marill et al. 2004, op. cit.).

The function of the ‘Frontier’ is to track URIs scheduled for harvesting. It is responsible for selecting the next URI to be tried and prevents the redundant rescheduling of already-scheduled URIs. The function of the ‘Processor Chains’ include modular ‘Processors’ that perform specific, ordered instructions on targeted URI in turn. These include fetching the URI, analysing the returned results, and passing discovered URIs back to the Frontier (Mohr, 2004, op. cit.).

4.7.1.2 LOCKSS

LOCKSS is a short for ‘Lots of Copies Keep Stuff Safe’. It is an open system model of physical document and it is applied for Web-published academic journals providing tools for libraries to take custody of the content of which they subscribe to and co-operate with libraries to preserve and enable access to such content (Maniatis et al., 2005).
LOCKSS preserves content that appears on a regular schedule and that is delivered through HTTP and have a URL. Content of Web sites that change frequently is not suited for archiving with LOCKSS. If a journal contains advertisements that change, such advertisements will not be preserved. Currently, it is being investigated if LOCKSS can be used to archive government documents published on the Web (Prudlo, 2005).

For improving security of content, LOCKSS runs from a CD rather than the hard disk (LOCKSS Website).  

4.7.1.3 NEDLIB

This crawler is an open source developed by a European Consortium since September 2002 at the Centre for Scientific Computing in Finland and runs on different UNIX platforms and relies on relational database (MySQL) for its configuration and process control. Its specification is written by NEDLIB and it supports the HTTP and FTP environment. It consists of a number of interrelated daemons such as scheduler, harvester, linkparser, linkfilter, performance mon, reqloader, doorman, metparser, and archiver. The National Library of Latvia has implemented the NEDLIB harvester (Marill et al. 2004, op. cit.; Wiggins, 2005, op. cit.).

---

81 LOCKSS open system accessed on 21/11/2007 from http://www.lockss.org/lockss/Installation_Instructions#Background
82 Accessible at http://www.csc.fi/sovellus/nedlib/
4.7.1.4 HTTrack\textsuperscript{83}

This is an easy to configure desktop crawler widely used and available for both Windows and UNIX systems. It is best suited for an explanatory acquisition of a small number of Websites. It modifies the links in retrieved content to create a self-consistent set of files that can be directly viewed without the need of a separate viewing tool. HTTrack is valuable for site analysis but not suitable for wide-scale harvesting (Marill \textit{et al.} 2004, \textit{op. cit.}). The German National Library develops and implements this crawling system (Wiggins, 2005, \textit{op. cit.}). Both Heritrix and NEDLIB require a separate tool to handle the presentation of the harvested content. The Internet Archive’s Wayback Machine, a proprietary presentation tool, is used for Alexa Internet and Heritrix output. While the NWA Toolset, released on December 2003, is an open source presentation tool designed to display the output of the NEDLIB harvester, (Marill \textit{et. al} 2004, \textit{op. cit.})

4.7.1.5 DeepArc

This crawler is developed at the BnF in 2003 for an experiment on voluntary deposit. It is discussed by the IIPC and accepted as an element of the toolset. It is available for test as free license software on Source Forge. The tool offers a graphic user interface to write the mapping between the database and a target XML Schema and the migration of the database into a flat XML file appropriate for preservation and possible reuse in the archive access context (Lupovici, 2005)\textsuperscript{84}.

\textsuperscript{83} Accessible at http://www.httrack.com
\textsuperscript{84} Lupovici, Catherine (2005). Op cit, p. 5.
Web crawlers and spiders make link checking and site archiving easy and are absolutely essential for search engines. A striking feature of Web crawling is that it depends on the interconnectedness of the Web, i.e. its link structure. This characteristic makes it difficult for crawlers to harvest restricted deep content of Websites and databases (Parsia, 2003, op. cit.).

Herein lies part of the challenges and problems of the current study are considered. This is because restriction often applied to prevent unauthorised access to content deemed to be for purchase. Therefore, unauthorised access linked to commercial content is regarded as an obvious infringement of IPRs on the networked environment. However, there are many other crawlers systems such as the WERA, ARC, and NutchWAX (Mohr, 2006, op. cit.) and Hanzoweb among many others (Middleton, 2006).

4.7.1.6 Wayback

This software operates as an access tool for navigation of Web archives. The software is called Wayback Machine to refer to a device used by professor Peabody in the ‘Rocky and Bullwinkle Show’ to travel through time. It is developed in 2001 by collaboration between Alexa Internet and the Internet Archive to provide public access to the full content maintained by the Internet Archive. It consists of four primary components, each with several implementations that can be combined to customise installations of varying scales and capabilities. The four components are query UI\textsuperscript{85}, resource store, resource index and reply UI (Tofel, 2007).

\textsuperscript{85} User Index.
4.7.2 Applying an Intranet System

The intranet contains a computer network connected to a number of corporate clients via the TCP/IP and hypertext transfer protocol. It is described as a private Website that provides proprietary services for certain organisation or group of internal local area network. It also looked at as an IP-based network of nodes installed behind firewall (Swantek, 2003,p.1418).

Intranets are built with the same tools that are used to build Websites. Both intranet and the Web require the same network protocol (TCP/IP) and they use e-mail and Web standards. The popularity of intranets in organisations has followed the growth of the Web. Many intranets are extensions of the National Library’s public Website and provide more in-depth information or proprietary information that would not be available to those outside the Library. Intranets are also being developed by groups of independent individuals of any organisation but with a common interest or mission. This instance of an intranet would provide a secure forum for discussion and information and knowledge sharing for those with common intranets (Ibid).
4.7.3 Applying an Integrated Archiving System

The proposed Sudan Web Archive could rely on the latest versions of the PANDAS archiving system of the National Library of Australia. This is because it supports the following preservation functions:

- Manage the metadata about titles that have been both selected and rejected for inclusion in the archive;
- Initiate collecting of titles selected for archiving;
- Manage the quality checking and problem fixing process;
- Prepare items for public display and generate a title entry page;
- Manage access restrictions; and
- Provide management reports (Phillips, 2003).

PANDAS software is a Web-based and enables all other library partners from their remote locations to conduct all of the tasks necessary to download and store titles in the central archive located on the National Library’s server. The system runs on a standard personal computer and with Internet Explorer 5.5 or advance (Philips, 2004, op. cit.).
4.7.4 Storage Technologies

A database needs to be created to store and retrieve the content. According to ISO, estimates for cost of storage is USD$5-USD$7 per megabyte when it includes items like labour (ISO, 1995, op. cit). Jantz (2003) claims that storage technologies cost USD$1.00 per gigabyte. While other researches (Michele, 2004, op. cit.) provide that the storage capacity costs USD$3,000 per terabyte. Units of the digital storage measurement are Kilobyte, megabyte, gigabyte, terabyte, petabyte, exabyte, zettabyte, and yottabyte as detailed in appendix (E).

Digital preservation systems will be more robust and have less probability for data corruption or loss if their storage component is a preservation aware storage, namely if their storage has built-in support for preservation. As discussed in the literature review section (1.9.1.1) the OAIS system is an ISO standard that specifies how digital assets should be preserved for a community of users. This system assist in defining requirements of an OAIS-based preservation aware storage as follows:

- Encapsulate and physically co-locate in the storage the raw data and its complex interrelated metadata objects, such as representation information, provenance and fixity. This ensures that the metadata needed for interpretation is not separated from the raw data and thus never lost (if the raw data survives).
- Utilise the local property and execute data intensive functions such as fixity computations within the storage component.
- Include the representation information of metadata such as the representation information of fixity and provenance, so
that the metadata can be validated and interpreted when migrating to newer systems.

- Handle the provenance events internally. The applications on top of the preservation aware storage should be freed from managing events that can be handled internally in the storage. Moreover, the types of provenance events are richer and also include events related to migration and transformation.

- Supporting the loading and execution of external transformations during the migration process. Additionally, it should facilitate on demand triggering of those transformations.

- Support media migration, as opposed to system migration in which migration from one system to another can be done by physical detaching the media from one system and attaching it to the new system.

- Maintain referential integrity including updating all the links during the migration process in order to remain valid in the new system. This requires an awareness of certain metadata fields that represent links, both internally to the system and externally.

- Ensure readability of the data by a different system in the future. This is done through developing and supporting global self-described media independent formats.

- Support a graceful loss of data. Some portions of the data are likely to be lost or become corrupted over time. If some data is lost, a good preservation system should minimize the economic effect of this data loss and prevent cases where
data in the system that is still intact cannot be read or interpreted (Lavoie, 2004).

Recent technological advancements made it easier for Web archive to index and control their holdings. For instance, the Storage Resource Broker (SRB) data grid is used to index and archive Web material (Minor, et al., 2007, p.3).

As mentioned in the introduction of this study, in the previous proposal, the researcher suggested use of the PDF format in digitisation of library materials in order to cope with deterioration of storage media. In the same time of submission of the proposal to the University of Khartoum, the PDF is adopted as ISO file format standard known as PDF Archive (PDF/A-1) (ISO 19005-1, 2005). It aims to represent documents, which are created natively in PDF form, converted from other electronic formats or digitised, from paper, microform, or other hard copy format to retain their quality and integrity. In June 2008, a new draft standard is submitted to ISO for approval known as WARC (Web ARChive) file format (ISO/DI 28500 Information and documentation—The WARC File Format). It is expected to function as standard method to structure, manage and store billions of resources harvested from the Web. It is an extension of the ARC file format that is used to store Web crawls as sequences of content blocks captured from the Web. The Internet Archive invented and implemented this format and used by other national libraries since 1996. Such extension request arose from the discussion and experience of the IIPC (ISO, 2008).
4.7.5 Creating the Metadata

The standard definition of metadata is always has been data about data (ALA, 2004-2006). Also metadata is viewed as structured information used to locate, access, use and manage information resources primarily in a digital environment (Fietzer, 2002,p.80).

Another approach to the term ‘metadata’ is that it is structured information which describe, explain, locate, or ease retrieval, use, or management of an information resource (NISO, 2004).

The metadata world hosts a range of systems to which the label of ‘metadata vocabulary’ is applied. These vocabularies range from basic descriptive metadata systems, with limited or single-focused functionalities, to more complex ‘member or class’ semantic vocabularies. An example of a simple ontology is the Dublin Core Metadata Initiative (DCMI) Elements and Element Refinements, a metadata schema developed mainly to facilitate resource discovery (Greenberg et al. 2003).

A more complex metadata vocabulary to which the word ‘ontology’ is applied is the Ariadne Genomics ontology. This ontology is used to formalise data on cell proteins for computer analysis. This ontology defines the various proteins, classifies them into taxonomic trees and defines the semantic relationships among them.

While many metadata vocabularies in operation were not necessarily created with the Semantic Web in mind, they may be able to play a significant role in its development. For existing and developing ontology to be used and function fully in the semantic Web environment, they need to adhere to

---

86 Accessible at www.dublincore.org/usage/terms/dc/current-elements/
87 Accessible at www.ariadnegenomics.com/technology/ontology.html
standards supported by enabling technologies (NISO, 2004, op. cit.).

The metadata structure is a fundamental component of the semantic Web. The Semantic Web has been envisioned by Tim Berners-Lee and is now being further defined by researchers and visionaries, but it was inspired by a host of creative thinkers who have, throughout history, looked to technological innovation as a way not only to control, but also to transform the world’s mass of information into intelligence (Greenberg et al. 2003, op. cit.)

Tagged (labelled) content enables the crawler to know what the content means semantically. This is the new idea envisioned as the ‘Semantic Web’ (Cerf, 2005, op. cit., pp. 19,21). This idea is developed through researches on hyperlinks of documents that connects different forms of data and allows computers to interlink and exchange these data on a networked platform (Berners-Lee, 2000).

The Semantic Web is rooted in Tim Berners-Lee’s original conception of the Web. It emphasises making Web-based content friendlier for programmes, especially reasoning programmes. Web links, however, have been seen not just as providing a navigatory connection for the reader, but also as partially constituting the meaning of the linked representations. In this sense, the Web is a kind of ‘semantic net’ with Web pages as nodes and hyperlinks as arc (Parsia, 2003). Metadata works to enable preservation managers to take appropriate action to preserve a digital object’s bit stream over the long term, and to ensure that the content of the archived object can be rendered and interpreted. But without data security, preservation is compromised.

OCLC’s digital preservation project for building Digital Archive infrastructure, metadata, and processes began in January 2001. The project has
three major goals which are to build a general-purpose digital archive for libraries, archives, and museums that may be used to store a variety of types of information and upon which various products and services may be built, to identify workflows for capturing and managing digital objects; and to implement a metadata set for the archived objects. When the first phase is completed in May 2002, the system has facilitated the capture of Web documents, creation of preservation metadata for digital objects, ingestion of objects into the digital archive, and long-term retention of these digital information assets. The first objects in the OCLC Digital Archive will be born-digital content and mostly public domain government documents published on the Web and consisting of text and still images presented in HTML, PDF, JPEG, GIF, BMP, TIFF, and ASCII formats (Bellinger, 2002).

The OAIS reference model [ISO 14721:2002] grew out of the need of NASA\textsuperscript{88} and other national space agencies to capture, access, and store vast quantities of digital information for the long-term as referred to in Chapter (1), section (1.9.1.1). While the details of the reference model are complex, the overall concept is a straightforward sequence of input-process-output (loc cit., p. 45).

National bibliographies need to encompass metadata relating to digital resources that may contain additional content (digital preservation data, terms and conditions of use) not required for print publications. New formats have been developed to contain such data and the traditional MARC formats have been updated to accommodate at least some of this additional information (Day et al. 1999, op. cit.).

\textsuperscript{88} National Aeronautics and Space Administration (NASA)
The metadata is a key to ensuring that resources will survive and continue to be accessible in the future. There are many different metadata schemas have been developed in a variety of user interfaces and disciplines. Some of the most common metadata schemas are Dublin Core, the Text Encoding Initiative (TEI), Metadata Encoding and Transmission Standard (METS), Metadata Objects Description Schema (MODS), Encoded Archival Description (EAD), and Learning Object Metadata (LOM). For its importance, only the Dublin Core Metadata will be discussed herein. With an emerging international standard for networked digital resources, primarily available via the Web. This standard is known as the Dublin Core and it provides guidelines for the metadata. It is intended to facilitate the discovery of digital resources and originally conceived for author-generated descriptions of Web resources. It represents one of many guidelines have been discussed in the area of digital resources. In the context of the Dublin Core, metadata includes what is generally referred to as bibliographic description whose elements may be used to find, identify, select and obtain resources available on the Web. Significantly, the Dublin Core includes fifteen data elements that constitute a core or basic form of description. It also discusses the critical issue of relationships to other works, expressions and manifestations, much in the same way as the IFLA standard does (Madison, 1998).

Compared to IFLA standards, fourteen of the fifteen of the Dublin Core elements are 'loosely' in common with the recommended IFLA data elements. The only Dublin Core element missing in the IFLA standard is the specific element of 'resource type.' The 'resource type' represents a textual description of the content of the resource or content description in the case of visual
resources (e.g., poem, dictionary, or musical recording). While the IFLA study group recognized its importance to the select function as an attribute of an expression, hence it gave it a low core value. In the future, as the pertinent IFLA standards are reviewed in terms of format specificity, the inclusion of this element might represent potential revision (Ibid.).

In any adopted digital preservation strategy, preservation metadata is likely to be a key part of its implementation. Lynch (2001, p.162) describes the function of some of this metadata that accompanies and makes reference to each digital object and provides associated descriptive, structural, administrative, rights management, and other characteristics.

These functions provide a clue that preservation metadata should enable to do more than support the implementation of any particular preservation strategy. Day (2003, op. cit.) suggests, for example, that metadata could be used to help ensure the authenticity of digital objects, to manage user access based on IPRs information as well as for more traditional metadata applications like resource description and discovery. Dollar (1998) emphasises that the preservation requires that records (documents) should be more than readable and intelligible, while those preserved for future should be identifiable, encapsulated, retrievable, reconstructable, understandable, and authentic. There are three types of preservation metadata: discovery, administrative, and structural metadata (Cerf, 2005, op. cit., p.39).

Often considered a type of administrative metadata, preservation metadata is coming into its own and has received more attention in the literature recently. Knight (2005) discusses the experience of the National Library of New Zealand (NLNZ) and their early work on the development of preservation metadata. He raises interesting
questions, such as what role will automation play in the creation of preservation metadata, when should it be captured, who is doing the updating, and how it should be done.

An important reason for creating descriptive metadata is to facilitate discovery of relevant information. In addition to resource discovery, metadata can help organise digital resources, facilitate interoperability and legacy resource integration; provide digital identification, and support archiving and preservation (NISO, 2004, op. cit. p.2).

Lavoie and Gartner (2005) provide a very readable overview of preservation metadata and why it is critical for digital objects, as well as a history of the development of a shared preservation element set and data dictionary called Preservation Metadata Implementation Strategies or PREMIS.

Caplan and Guenther (2005), co-chairs of PREMIS, the OCLC/RLG Working Group on Preservation Metadata Implementation Strategies, provide greater detail on the creation of PREMIS, which comprises basic preservation metadata elements tied to an XML-data dictionary to avoid being platform specific. With digital objects and digital projects growing exponentially, it becomes even more critical to work on the total framework and the management of the metadata being created (Caplan & Guenther, 2005).

The NLNZ schema identifies the data that the Library will collect and maintain Web resources. This relates to the preservation master held in the Digital Archive, but could also cater for an object that is not or is no longer a preservation master, e.g., the CD-ROM on which the Library received the original digital object or a previous preservation master that has been superseded through hardware or software obsolescence (Searle & Thompson, 2003).
Such preservation master will be a ‘best effort’ creation of a working preservation object. It will be a rendition of some form of ‘original’ as supplied to or acquired by the Library in a file format that can be preserved, managed and disseminated over time. The preservation master is dynamic and will be subject to processes such as migration and emulation during a lifecycle of creation, use and eventual replacement. At any time there can be only one preservation master for an object and maximum preservation effort will be applied whilst it has the same status. As shown in the diagram below, the NLNZ schema is split into four entities (Knight, 2005, op. cit.).
Figure (4.3) The four entities of the NLNZ metadata schema. (Source: Searle & Thompson, 2003, op. cit.)
4.7.6 Creating the national bibliography

One of the main objectives of the proposed SUWA is to maintain the national Web bibliography. The bibliography is a science of standard listing and analytical study of books, manuscripts and other print library material (Feather & Sturges, 1997, p.30).

It is usually complied with the intention of providing comprehensive coverage of its chosen field. This field might be defined chronologically, geographically, by subject, by author or by format of publication, or by some combination of one or more of these. Such form of bibliography is known as enumerative bibliography, a term which encompasses almost all bibliographies (Greetham, 1994). There are three types of bibliography, which are commonly distinguished as author bibliographies, subject bibliographies, and national bibliographies (Feather & Sturges, 1997, op. cit.).

However, bibliography is closely associated with bibliometrics. A term first appeared in 1969 to replace ‘statistical bibliography’. It refers to the study of the use of documents and patterns of publication, by means of mathematical and statistical methods. Bibliometrics can be divided into ‘descriptive’ and ‘evaluative’ count both of which can, in turn, be further divided by ‘productive count’ (geography, time and discipline) and ‘literature count’ (reference and citation). There are also two other terms, used more or less interchangeably, which are scientometrics and informetrics (Hertzel, 2003, p.288), (cf. Chapter (1), section (1.9.1) and Chapter (3), 3.1).

National bibliographies are a record of the publications of a country and or works by authors who are resident and or born in and or ethnic nationals of the country. In the wider sense this
means the listing of information products relating to a particular nation. Practically, national bibliography encompasses a diverse range of materials such as books, pamphlets but also deals with the whole national documentary heritage other than which is properly the domain of the national archives. Ideally its scope should include serials, printed music, maps, audiovisuals, government publications, and digital information products. In most cases this means materials produced within the national territory, but is often broadened to include material on the subject of the nation, or some specific feature of the nation; material produced by members of the nation, originating whosessoever it might, material in the national language (Feather & Sturges, 1997, op. cit., p. 312).

There are three main goals have been traditionally assigned to national bibliographic services. The first is to assist cost-effective cataloguing in libraries. The second is to assist libraries in the selection and acquisition activities. The third objective is to further information searching and retrieval for document supply. These objectives are still valid. What is changing today is the environment in which such objectives are implemented (Vitiello, 1998).

National bibliography, however, is generally subdivided into retrospective bibliography and current bibliography. Retrospective national bibliography has usually been achieved by the publication of the catalogue of the national library, usually relying on the institution’s collections, which result from the exercise of legal deposit privileges. The published bibliography of the Library of Congress exemplifies this type (Feather & Sturges, 1997, op. cit.).

The book trade originally undertakes current national bibliographic work as an aid to commerce
in books and this activity still survives in publications like the Cumulative Book Index in the US. Since the UNESCO Conference on the Improvement of Bibliographic Services in 1951, regular listing of books and other documents has been seen as a function of a national agency, often but not exclusively forming part of the national library. The degree of currency actually achieved by many national bibliographies, which are published, varies greatly from country to country, but long time lags between the appearance of a document and the appearance of the listing are the rule rather than the exception (UNESCO, 1961).

The development of national bibliographies is connected to the Universal Bibliographic Control (UBC) standards. The purpose of the UBC is identified as for utilising of and exchanging international bibliographic records created in certain country though based on international bibliographic standards (Day et al. 1999, op. cit.) This also based on the premise that cataloguers in any one country are best able to describe the publications of their country. The UBC presupposed the creation of systems that could be used for the international exchange of standard bibliographical descriptions of publications such as the International Standard Bibliographic Description (ISBD) and the UNIMARC format.

The French Web Legal Deposit is the first law to include preparation of Web-based legal deposit. The result will be a list of URLs with name index, subject index (as far as it is possible to make it with a linguistic analysis, close to what has been done in the Watson project (Lupovici, 2005, op. cit., p.5).

Nowadays, the aim is to create not simply a serial bibliography published on paper, but a digital national bibliographic database in MARC format. Such a database could be used to generate bibliographic products in paper or could be formed
and made available online via services such as BLAISE-LINE (Feather and Sturges, 1997, op. cit.).

However, the proposed Sudan Web Archive should be responsible for creating a comprehensive national bibliography published on the ‘.sd’ domain and beyond. It is easy to create a national bibliographic database using WorldCat catalogue and search other catalogues of libraries, which are published on the Web in order to complement the defects in coverage.

4.7.7 Implementation of the plan

The plan should be implemented as quickly as possible. The longer it takes to implement the strategy, the less of a chance it will be successful. Focus on Web-based intranet applications that have revenue generation potential as well as cost-saving benefits. The design of the Sudan National Library’s Website, intranet and databases (technology) should reflect its processes rather than simply representing its organisational chart. This is because the Web Archive is an inseparable part of the Library services. Therefore the library could be able to support the communication of various data types within its own infrastructure.

Using a consistent design format for the presentation of all information makes it easy for staff and users to navigate the library Website and intranet and obtain information and services. The interfaces for all pages should be based on a common theme. Site navigation should be intuitive and consistent. Ideally, the Websites should have an easy to follow, intuitive navigation schema that provides users with both search and download facilities. These sites should use icons or other visual and organisational metaphors as navigation tools, or use simple tree structure to organise the Website. To determine the best design for the
Library Website, input from the potential user groups is critical. Design teams should obtain input and ideas from the intended audience in focus groups and consider views of the librarians and other information professionals. Key issues to be addressed when implementing the preservation plan are to:

- Define the mission and purpose of the Web Archive including its structure, divisions, subsidiaries, and departments. Cultural, geographical and linguistic issues should be considered.
- Review the technology platform in place within the National Library mission and vision and determine the management of the technology.
- Establish the process for content development and management, what will be included, how often to update, licencing or permission for external content, privacy, security guidelines, and multilingual other issues.
- The simple applications should be launched first, implementing them in an area that has the most influence on the rest of the Library services.

The preservation application should be piloted with a small test group before rolling it out to the larger audience. This allows testing the technology performance, determining future time and costs required to maintain the services of the Archive, and collecting additional feedback and user input. The final guidelines for establishment of the Sudan Web Archive could be provided as follows:

4.7.7.1 Infrastructure

Hardware components such as high-speed and greater computer memory, servers, Internet connection devices should be provided. A
Website for the Archive should be launched within the hosting national library or as separate one. Software such as digital archiving management system, antivirus, crawler, file format and others need to be provided.

4.7.7.2 The Selection Plan

The seed list based on the '.sd' domain and related foreign domains need to be prepared. Also identification of the capturing and harvesting scope is essential. Quality of content, frequency of capture and user groups, information literacy are key to achieving success.

4.7.7.3 Legality of the Preserved Content

In accordance with the proposed Web deposit the Archive should commence with capturing the content which have fallen into the public domain, official governmental publications, daily news or occurrences of published newspapers, magazines, or other periodicals or broadcast of radio or television, ideas, methods, state emblems and symbols, traditional knowledge and folklore.

4.7.7.4 Proprietary Content

Since there is no clear authorisation in the Sudan IPRs legislation, proprietary content should be preserved after amending the related laws and or obtain such content through licencing at this initial phase. Other proprietary content relevant to the country should be negotiated with their rightholders for harvest.
4.7.7.5 The Descriptive Metadata

The metadata should identify level of description, and Website level description would by sufficient metadata elements should contain: author, title, date, URL address, and other subjects as well as building natural language vocabulary for indexing and retrieval.

4.7.7.6 Display and Access

The Archive search facilities should enable the discovery of items stored in it via search screen based on knowledge of user practices in search engines. Subject categorisation with drop-down menus and lists would be useful if provided. Access should be unrestricted to users pursuant to the right to information principles. The look-and-feel features should be retained during implementing the preservation strategies. Also it is not necessary to enable links to dead pages not included in the collection. While it would be useful to maintain URLs to dead pages not included in the collection and ensure authenticity of items displayed.

4.7.7.7 Maintenance and Weeding

Maintenance of links to changed URLs is encouraged if possible and the disappeared URLs should be deselected.
4.7.7.8 User Feedback

A platform for direct feedback should be enabled to enable interaction with users via e-mail messages such as ‘Ask A Librarian’ facility.

4.7.7.9 Preservation

It is feasible to migrate, refresh and emulate hardware devices, which operate content when such devices become obsolete. Migration of content into new file format is essential for long-term access.

4.7.7.10 Implementation of the Strategies

In order to ensure longevity of access to the preserved content, implementation of the preservation strategies should be implemented in order to ensure longevity of the life cycle of the holdings to the future Library patrons.

4.7.7.11 Duration

Long-term storage of holdings is defined in this study as longer than the lifetime of the hardware and software or as long as required as stated in the WARC draft standard (ISO, 2008, op.ct.).

4.7.7.12 Budgets

The financial costs of storage, preservation and maintaining access to content should be allocated and made ready as part of the digital national library project.
4.7.7.13 Staffing

At the initial phase, two subject librarians and a Web archive manager could co-operate with the exiting information technology staff to build the collection and set the scene for future job descriptions. A consortium of information institutions is recommended.

4.7.7.14 Security and Integrity

Access authentication methods should be enabled to protect unauthorised access to the administrative data entry and processing interfaces. Pre-emptive measures should be taken in order to predict network failure, power shortages, virus attacks, destruction, and deliberate damage trials.

Precautions should be dealt with when selecting Web-based content for long-term preservation. Web-based content, according to Kenney et al. (2002, op. cit.) is particularly vulnerable to external attacks. In 2001, for instance, the Nimda worm, which has taken down 150,000 computers, and the Code Red Virus, which struck more than 12,000 Websites in the US, hit the Web hard. In June in the same year, Microsoft has issued a patch to protect its attack-vulnerable Internet Information Server (IIS) software, which is used by approximately 16 million Websites.

As a consequence, a risk management strategy should be associated with content filtering in order to protect the Archive from viruses, piracy and hacking attacks.
4.7.7.15 Disaster Recovery

To maintain the integrity of the storage and retrieval for long time enough, a disaster and emergency plan need to be put in detail. A precise sketch should be developed for disaster forecast and risk management programme.

4.7.7.16 International Co-operation

Building international relationships is vital for acquiring knowledge and experience on Web archiving operations. Also attendance at workshops and conferences is highly desirable so that staff could exchanging skills and expertise with the successful projects in order to share views for common standards and troubleshooting.

This plan should be provided to the decision-makers at the Sudan National Library for review and action. The Board of the National Library comprises a well-qualified library and information professionals who can approve this proposal.

Summary

It is clear that building SUWA is a matter of time. As long as the above plan is delayed, great cultural and scientific heritage will be lost. The revised 1971 Legal Deposit Act needs to be updated and Web legal deposit should be enacted to enable establishment of the Web Archive. The next Chapter aims to interpret the findings and results of the survey and concludes the study.
CHAPTER 5
THE CONCLUSION

Overview

This is the last Chapter of the study, which aims to provide interpretation of findings, discussion of the outcome and the conclusion connecting all the previous Chapters together.

5.1 Interpretation of Findings

Procedures followed in the methodology Chapter have been successful in collecting and analysing the results. The questionnaires are forwarded in the period between March through May 2007 as purposive e-mail attachments to targeted sample of staff of national libraries and archives of the world represented on the Web. Total responses received are 1200 from 136 institutions worldwide after three reminders. There is no missing data recorded. There are 10 institutions reported e-mail failure demon. A response rate is 11.00%, which is good enough in comparison to the newness of the undertaken topic considering that the Web itself is in its teenage. While the decline rate is high at 23.00% and this might be attributed to many reasons including most national libraries and archives do not answer e-mail attachments, e-mails posted on Websites of these institutions are no longer updated, failure demons received as a result of technical problems of the receivers, or due to other unknown reasons.

Generally, the survey is analysed and its results are reported in Tables (5.1-5.4) and presented in Figures (5.1-5.10) as shown in appendix (A) at the end of the study. Results of analysis of variance (ANOVA) are presented in Figure (5.5), correlations in Figure (5.6), ratio statistics in Figure (5.7), regression in Figure (5.8), reliability in Figure
(5.9) and T-Test in Figure (5.10) all show that there is strong and significant relationship between the variables of the study as discussed in Chapter (1) sections (1.9.3 and 1.9.4) and the outcomes of section (2.4) in Chapter (2). The quantitative statistical data reached through procedures followed in Chapter (3) concluded that the correlation between the importance of preservation of Web-based content and the barriers of IPRs (QuestionB and QuestionC) is significant at (0.05 level) as indicated in Figure (5.2). Such correlation could possibly be linked to the finding of the first dependent variable (refer to Chapter (1) section (1.9.3) which indicates that though 7 million Web pages are added on daily basis as Layman (2002, op. cit.) reports but Web-based content disappear at greater rates. The Library of Congress Report (2002, op. cit.) provided a fact that could be generalised that annual Web-based content disappearance rate is 44% and average life of a Website is 44 days or 75 days as of Day (2005, op. cit.) or 25% as shown in Table (5.5) and Figure (5.2) in appendix (A).

Since $r$ (correlations) = -0.05 (cf. Figure 5.6) is negative and a little closer to -1 than to 0, the study would tend to conclude that there appears to be negative linear relationship between the right to copy for preservation in national libraries and the IPRs restrictions - that is, as IPRs restrictions increase on the Web, as copying limitations tend to decrease. Therefore, the hypothesis, which states that ‘Web-based content is ephemeral, fragile and prone to loss due to Websites change and disappearance’ as suggested in section (1.3.1) is proved valid. Similarly, for the hypothesis in section (1.3.2), which states that ‘the proposed Sudan Web Archive cannot preserve the Web-based content unless both strategies and legislation are prepared’, is proved valid for lack of legal authorisation as the interview data
reveals. The hypothesis in section (1.3.3), which states that 'Copying for preservation of Web-based content will be restricted because of recent trends of the GATS and TRIPS agreements towards its commercialisation', is found valid and since the Sudan conformity to the WTO is unapproved yet this could give chance to share views with the National Library professionals to include exemptions for preservation. The hypothesis in section (1.3.4), which provides that 'Preservation of Web-based content is vulnerable to terrorists' attacks, Internet backbone destruction, viruses, hacking and piracy' is proved valid as supported by the Internet Archive in section (1.9.1.1.3) and Chapter (2), section (2.4.5). The hypothesis in section (1.3.5), which states that 'preservation of Web-based content including proprietary software, hardware, databases and other surface Web-based content cannot be done due to restrictions of intellectual property rights', is valid as induced from hypothesis in section (1.3.1) as proved above. The last hypothesis in section (1.3.6), which provides that 'preservation of Web-based content is affected by the lifecycle of obsolescence and dynamic change of software and hardware' is also proved valid as induced in section (1.3.6) as above-mentioned.

Nonetheless, the interviews with custodial administrators in the Sudan provided no additional quantitative figures to measure the problems they face and their perceptions of various policies concerning digital preservation generally. Probably, this may be attributed to their unawareness of Web-based content as cultural heritage as stated in the problem of the study and as discussed in the literature review section (1.9.3).

An interview was made with Mr Ibrahim Al Ameen, Acting-National Librarian of the Sudan, revealed that he and his staff were unaware of the
importance of preservation of Web-based content for future access as national heritage (Al Ameen, 2007).

An interview also was conducted with Ms Najwa Mahmoud who is in charge of the digitisation project at the NRO but she provided that she and her staff were unaware of the Web-based content as essential part of the national digital heritage and as part of the NRO collections (Mahmoud, 2007).

5.2 Discussion

It is clear that the annual rates of content disappearance is high and the short life of the Website urge the Sudan National Library to adopt establishment of the Web archive to maintain access to Web-based content as proposed. An important finding reached is that great deal of the Sudan Web-based heritage probably disappeared during cessation of the country’s domain in the period from mid 2001 until 18 November 2002 as discussed in Chapter (4), section (4.1). Findings also reveal that there are major challenges to allow comprehensive Internet penetration in the country, which may affect access and utilisation of the Archive’s holdings. As defined in Chapter (1) section (1.8.1), the proposed Web Archive is ready to be presented to the decision makers at the Sudan National Library Board in order to take an immediate action towards affording funds and facilities to build its collection before it is too late.

5.3 Limitations of the Study

It is a fact that no research is complete and perfect. There are points of weaknesses in this study. In particular, no details on how to tackle the link boundaries are made. Therefore, the study has unveiled fertile areas for several subtopics
worth for future digital preservation research works as follows:

- Citation analysis of Web links;
- File format for Web archiving;
- Consumers and Content quality on the Web;
- Integrity and safety of Web content for Web archive users;
- Privacy of information in Web archiving;
- Web archiving and challenges of Piracy and hacking;
- Web legal deposit as a collection development tool;
- LIS e-journals: a Webometric approach;
- Training and capacity building for webometricians;
- Presence of academic OPACs on the .sd domain;
- Impacts of Internet penetration on access to Web archive;

**Conclusion**

It is interesting that the content of the literature review has made the researcher learn more about the nature of this kind of study. The most interesting is that the former proposal he has suggested for this study has been adopted as an international standard ISO PDF/A as mentioned in the introduction. This research is a useful voyage of discovery for building better skills and relationships between the supervisors, fellow librarians and information professionals worldwide. All the objectives of the study are met. The study identified various research areas that could be carried out for further investigations. This type of research worth doing because it advances knowledge on LIS curriculum including contribution to knowledge on Web archiving, filling gaps of literature in this field, building the Sudan Web
Archive and to the emerging Webometrics field within LIS. Finally, the study is not meant to be perfect but it is an endeavour for making a success.
المراجع

الخطط والتقارير

دار الوثائق القومية، (د.ت.). كتيب تعريفي عن دار الوثائق القومية. أخرطوم: رئاسة مجلس الوزراء. صص. 2-3

حمد، رضية آدم، 1989. إنشاء بنك المعلومات برئاسة مجلس الوزراء. أخرطوم: مركز علوم المعلومات _ جامعة أخرطوم.

المركز القومي للمعلومات، 2006. التقرير النهائي للجنة تنسيق معلومات المكتبة الإلكترونية. أخرطوم: وزارة رئاسة مجلس الوزراء.

______________________________________________________________، 2006. خطة عمل المركز القومي للمعلومات. أخرطوم: وزارة رئاسة مجلس الوزراء.


القوانين والتشريعات

السودان. قانون حقوق المؤلف والحقوق المجاورة لسنة 1996م. أخرطوم: المجلس الإقليمي للمصنفات الأدبية والفنية.

السودان. قانون الصحافة والمطبوعات الصحافية لسنة 2004م. أخرطوم: المجلس القومي للصحافة والمطبوعات.

السودان. قانون المركز القومي للمعلومات لسنة 1999م. أخرطوم: رئاسة مجلس الوزراء.
السودان. قانون المصنفات الأدبية والفنية لسنة 2000م

السودان. قانون المكتبة الوطنية لسنة 1999م. الخرطوم: وزارة الثقافة والإعلام.


السودان. مشروع قانون مكافحة جرائم المعلومات لسنة المركز القومي للمعلومات. 2006م. الخرطوم:
REFERENCES

Books


Reports


OCLC, 2004. *Information Format Trends: Content, Not Containers*. Dublin, Ohio. Available at:


**Periodicals**


__________, 2005. Google Library Project Hit by Copyright Challenge from University Presses.


__________, et al., 2003. Linguistic Patterns of Academic Web Use in Western Europe,


**Conference Papers**


324


Jones, Maggie, 2003. Digital Preservation Activities in the U.K. - building the


http://nedlib.kb.nl/workshop/pat1.ppt [accessed 04 May, 2008]


Website Information


_____________. Available at: http://info.cern.ch/ [accessed 04 May 2004]


IIPC Website. Available at: http://netpreserve.org/about/mission.php [accessed 29 March 2007]

THE INTERNET ARCHIVE WEBSITE. Available at: http://www.archive.org/about/about.php [accessed 22 November 2006]


The Library of Congress Website. Available at: http://www.loc.gov/about/history.html [accessed 27 April 2008]

NARA WEB SITE. Available at: http://www.archives.gov/preservation/internal/ [accessed 13 March 2007]

NIC WEBSITE. Available at: http://www.dalil.sd/dalil_sd.php [accessed 14 April 2008]


Project Gutenberg Literary Archive Foundation Website. Available at: http://www.gutenberg.org/wiki/Main_Page [accessed 30 September 2007]


UK Web Archiving Consortium Website. Available at: http://info.webarchive.org.uk/copyright.html [accessed 07 November 2007]

Webopedia Website. Available at: http://www.webopedia.com/TERM/F/first_sale_doctrine.html [accessed 01 September 2007]

Legal Cases


The EU v. Microsoft [2007]. In the Illinois Business Law Journal. Available at: 
[accessed 06 July, 2007]

Available at: 
[accessed 17 April 2008]

Available at: 
http://w2.eff.org/IP/DMCA/US_v_Elcomsoft/us_v_elcomsoft_faq.html 
[accessed 17 April 2008]

Viacom et al v. YouTube et al [2008]. Available at: 
http://docs.justia.com/cases/federal/district-courts/new-york/nysdce/1:2007cv02103/302164/95/ 
[accessed 06 July, 2008]

Laws and legislation

Agreement on the Creation of the African Regional Industrial Property Organisation (ARIPO) 1976. 
Lusaka: ARIPO. Available at: 
[accessed 27 September 2007]

Agreement Revising the Bangui Agreement 1977 on the Creation of an African Intellectual Property Organisation (OAPI). Bangui: OAPI. Available at: 
[accessed 27 September 2007]

Australia National Library Act, 1960. Available at: 
[accessed 20 July 2008]
Bangui Agreement, 1977. Bangui: OAPI. Available at:
[accessed 14 April 2008]

Berne Convention, 1886. Geneva: WIPO. Available at:
[accessed 13 August 2007]

Convention on Biological Diversity 1993. Montreal:
CBD. Available at:
[accessed 09 July, 2008]

Convention on the Protection and Promotion of the
Diversity of Cultural Expressions 2005. Paris:
UNESCO. Available at:
http://unesdoc.unesco.org/images/0014/001429/14
2919e.pdf [accessed 23 December 2007]

Copyright Act 1976. Section 107, p.9, as amended
2007, Washington, D.C.: Library of Congress. Available at:
http://www.copyright.gov/title17/circ92.pdf
[accessed 08 October 2007]

Copyright Act No. (148) of 2003 as amended.
Canberra. Available at:
[accessed 23 August 2007]

the European Communities. Available at:
http://cyber.law.harvard.edu/property00/alternatives/directive.html
[accessed 10 July 2006]
Finland Legal Deposit, 1984. Motion pictures, films and videos in the Finnish Film Archive. Helsinki.


Standards and Codes


Dissertations and Theses


**Personal Communication**

Al-Ameen, Ibrahim, 2007. Discussion on Web archiving. [Interview], (personal communication conducted at 11.00 a.m., 23\textsuperscript{rd} December 2007).

Mahmound, Najwa, 2007. Discussion on Web archiving. [Interview], (personal communication conducted at 11.00 a.m., 29\textsuperscript{th} December 2007).
APPENDIX (A)
QUESTIONNAIRE FORM AND ANALYSIS

QUESTIONNAIRE

‘STRATEGIES FOR PRESERVATION OF WEB-BASED CONTENT: INTELLECTUAL PROPERTY BARRIERS TO BUILDING THE SUDAN WEB ARCHIVE’

A thesis proposal submitted to the Graduate College at the University of Khartoum for obtaining the degree of Doctor of Philosophy (PhD) in Library & Information Science, 2005 by Gafar Ibrahim.

Introduction:

The Web comprises various topics on many disciplines, sciences, enterprises, education and learning, entertainment, media and other components. But most of these topics are not valuable for the scientific research and decision-making purposes. This is due to the open nature of Web publishing in contrast to traditional restricted peer-reviewed publishing. Undeniably, there are several Web databases of academic institutions, publishers, scientific and cultural associations, international organisations, information and research centres which produce reliable content for educational, cultural and for research and development. There is a major challenge that Web content is prone to disappearance forever as a result of ongoing update, editing of the content in response to additions, or deletions or changing over to new URLs. Increasingly, rapid advancements in the development and obsolescence of software and hardware cause risk of content inaccessibility.

89 Librarian & Information Officer at GHD Middle East Operating Centre.
Consequently, there is a growing need for setting up strategies for preservation of significant Web-based content for future research generations.

---

How to Fill in the Questionnaire

**Note for the MS Word users:**

Please double click the appropriate grey box then a dialogue window will appear then select (Checked) then click (Ok). You can cancel selection by double clicking the box then select (Not Checked) then click (Ok). After completing the questions please save the file and send it as file attachment via email to Mr. Gafar Ibrahim at: shash_1112@yahoo.co.uk

**Note for those who use a hard copy:**

In case you prefer filling in the questionnaire by handwriting, please tick the appropriate box and send it by fax to Gafar Ibrahim via fax No.: (+974) 444 6127.

**Note:** data privacy of all participants shall be dealt with top confidentiality.

**A. Personal Information:**

| Name: (optional) | |  
| Position: | |  
| Institution: | |  

### B. Importance of Web Preservation

<table>
<thead>
<tr>
<th>Why preservation of Web-based content is significant?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1- Because it is ephemeral and has no print equivalent to preserve:</td>
</tr>
<tr>
<td>2- Because future generations need to access historical and cultural Web-based content of the past:</td>
</tr>
<tr>
<td>3- Because every aspects of our Web activities need documentation as future scientific &amp; legal evidence:</td>
</tr>
<tr>
<td>4- Because it is vital to preserve Web-based content to ensure prospective flow and success of operations of e-learning, e-governance, e-commerce etc.:</td>
</tr>
<tr>
<td>5- Because I anticipate a sudden disaster that may hamper functions of TCP/IP(^\text{90}) which may cause loss of Web-based content forever:</td>
</tr>
<tr>
<td>6- Because rapid change &amp; obsolescence of software &amp; hardware affect access to Web-based content in the long-term:</td>
</tr>
</tbody>
</table>

\(^{90}\text{Internet Protocol/Transmission Control Protocols}\)
### C. Legal Issues

In your country, legislation on preservation of digital information including Web-based content is enshrined in:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1- The legal deposit law:</td>
<td>Yes [ ] No [ ]</td>
</tr>
<tr>
<td>2- The copyright law:</td>
<td>Yes [ ] No [ ]</td>
</tr>
<tr>
<td>3- The national information policy</td>
<td>Yes [ ] No [ ]</td>
</tr>
<tr>
<td>4- The national legislation on the Internet:</td>
<td>Yes [ ] No [ ]</td>
</tr>
<tr>
<td>5- The information and telecommunication law:</td>
<td>Yes [ ] No [ ]</td>
</tr>
<tr>
<td>6- The Digital Millennium Act:</td>
<td>Yes [ ] No [ ]</td>
</tr>
<tr>
<td>7- Other (specify):</td>
<td></td>
</tr>
</tbody>
</table>

Copyright & other intellectual property rights are regarded as major obstacles for preservation of endangered Web-based content:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>8- Copyright &amp; other intellectual property rights are regarded as major obstacles for preservation of endangered Web-based content:</td>
<td>Agree [ ] Disagree [ ]</td>
</tr>
</tbody>
</table>

There is a need for revision of legal deposit laws and polices in order to include collection & preservation of Web-based memory:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9- There is a need for revision of legal deposit laws and polices in order to include collection &amp; preservation of Web-based memory.</td>
<td>Agree [ ] Disagree [ ]</td>
</tr>
</tbody>
</table>

National Libraries are appropriate institutions for reinforcing provisions for preservation of the digital memory including Web-based content:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10- National Libraries are appropriate institutions for reinforcing provisions for preservation of the digital memory including Web-based content.</td>
<td>Agree [ ] Disagree [ ]</td>
</tr>
</tbody>
</table>

In your legal systems, what terms that govern duration of copyrighted material?

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>11- In your legal systems, what terms that govern duration of copyrighted material?</td>
<td></td>
</tr>
</tbody>
</table>

Every agency which has a Web site should be obliged to deposit a copy of its new or periodically updated content to a central custodial institution:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>12- Every agency which has a Web site should be obliged to deposit a copy of its new or periodically updated content to a central custodial institution:</td>
<td>Agree [ ] Disagree [ ]</td>
</tr>
</tbody>
</table>

In your legal systems, what terms that govern fair use?

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>13- In your legal systems, what terms that govern fair use?</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1</td>
<td>In your institution, licensed access to online databases ensures preservation requirements:</td>
</tr>
<tr>
<td>2</td>
<td>Viruses, terrorist attacks, hacking, piracy &amp; networks failure threat continuity &amp; flow of Web-based content:</td>
</tr>
<tr>
<td>3</td>
<td>Preservation of Web-based content reinforces national security:</td>
</tr>
<tr>
<td>4</td>
<td>Your institution Web preservation project is completed &amp; provide services to users:</td>
</tr>
<tr>
<td>5</td>
<td>Web-based content need filtering and assessment before being preserved:</td>
</tr>
<tr>
<td>6</td>
<td>When did your institution begin digital preservation including Web preservation project?</td>
</tr>
<tr>
<td>7</td>
<td>Estimate the project budget of your institution for digital preservation including Web preservation:</td>
</tr>
<tr>
<td>8</td>
<td>Collection &amp; preservation of Web resources should be one of the national library responsibilities:</td>
</tr>
<tr>
<td>9</td>
<td>Internet governance &amp; administration would reinforce credibility, suitability and availability of Web content:</td>
</tr>
<tr>
<td>10</td>
<td>In your institution what file format is used for Web preservation?</td>
</tr>
<tr>
<td>11</td>
<td>If your answer to question No. (26) is (other; please specify):</td>
</tr>
<tr>
<td>12</td>
<td>If you disagree with question No. (27), what institution is carrying out the digital preservation including Web preservation project?</td>
</tr>
</tbody>
</table>

---

91 File format used internally by the Internet Archive.
92 Web ARCHive.
93 Portable Document Format – Archive.
13- In your institution, what harvesting system is used for collecting & capturing Web resources:
-----------------------------------------------------------------
-----
14- Please provide your own comments about this survey:
-----------------------------------------------------------------
-----
15- Would you like a copy of the survey results to be sent to you?
Please provide your e-mail address: -----------------------------
---------------
## Data Analysis

**Question No. (A) Personal data**

Table No. (5.1) type of institutions responded to question No (A).

<table>
<thead>
<tr>
<th>Type of Institution</th>
<th>Responses</th>
<th>Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Libraries</td>
<td>203</td>
<td>17%</td>
</tr>
<tr>
<td>National Archives</td>
<td>345</td>
<td>29%</td>
</tr>
<tr>
<td>Web Archives</td>
<td>473</td>
<td>39%</td>
</tr>
<tr>
<td>Decline</td>
<td>179</td>
<td>15%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1200</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

Figure No. (5.1) Response Rates by Type of Institutions.
Question No. (B) Importance of Web Preservation

Table No. (5.2) response rates of question No (B).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Question B</th>
<th>Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ephemerality</td>
<td>98</td>
<td>25%</td>
</tr>
<tr>
<td>Accessibility</td>
<td>92</td>
<td>23%</td>
</tr>
<tr>
<td>Documentary</td>
<td>88</td>
<td>22%</td>
</tr>
<tr>
<td>Interoperability</td>
<td>91</td>
<td>23%</td>
</tr>
<tr>
<td>Disasters</td>
<td>11</td>
<td>3%</td>
</tr>
<tr>
<td>Change</td>
<td>17</td>
<td>4%</td>
</tr>
<tr>
<td>Total</td>
<td>397</td>
<td>100%</td>
</tr>
</tbody>
</table>

Figure No. (5.2) Rates of Reasons for Web Archiving.
Question No. (C) Legality of Web archiving

Table No. (5.3) the response rates of question No (C).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Question C</th>
<th>Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legal deposit</td>
<td>47</td>
<td>13%</td>
</tr>
<tr>
<td>Copyright</td>
<td>23</td>
<td>6%</td>
</tr>
<tr>
<td>Information policy</td>
<td>130</td>
<td>36%</td>
</tr>
<tr>
<td>Internet law</td>
<td>70</td>
<td>19%</td>
</tr>
<tr>
<td>Telecom law</td>
<td>55</td>
<td>15%</td>
</tr>
<tr>
<td>DMA Act</td>
<td>38</td>
<td>10%</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>363</strong></td>
<td><strong>99%</strong></td>
</tr>
</tbody>
</table>

Figure No. (5.3) Rates of Preservation Laws.
Question No. (D) General Web preservation issues

Table No. (5.4) the response rates of question No (D).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Question D</th>
<th>Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database licencing</td>
<td>15</td>
<td>6%</td>
</tr>
<tr>
<td>Threats</td>
<td>56</td>
<td>21%</td>
</tr>
<tr>
<td>Security</td>
<td>7</td>
<td>3%</td>
</tr>
<tr>
<td>Operation</td>
<td>12</td>
<td>5%</td>
</tr>
<tr>
<td>Filtration</td>
<td>65</td>
<td>25%</td>
</tr>
<tr>
<td>Start of archiving</td>
<td>12</td>
<td>5%</td>
</tr>
<tr>
<td>Budgets</td>
<td>12</td>
<td>5%</td>
</tr>
<tr>
<td>Responsibility</td>
<td>55</td>
<td>21%</td>
</tr>
<tr>
<td>Internet governance</td>
<td>13</td>
<td>5%</td>
</tr>
<tr>
<td>File format</td>
<td>4</td>
<td>2%</td>
</tr>
<tr>
<td>Crawler</td>
<td>10</td>
<td>4%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>261</strong></td>
<td><strong>65%</strong></td>
</tr>
</tbody>
</table>

Figure No. (5.4) Rates of Web Preservation Issues.
Oneway

[DataSet2]

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>11928.000</td>
<td>6</td>
<td>2838.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Groups</td>
<td>618.000</td>
<td>4</td>
<td>154.500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>12546.000</td>
<td>10</td>
<td>1254.600</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>3497.392</td>
<td>6</td>
<td>582.897</td>
<td>1.382</td>
<td>.393</td>
</tr>
<tr>
<td>Within Groups</td>
<td>5688.800</td>
<td>4</td>
<td>1422.200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9186.192</td>
<td>10</td>
<td>918.619</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure No. (5.5) Analysis of Variance (ANOVA)
**Correlations**

<table>
<thead>
<tr>
<th></th>
<th>QuestionB</th>
<th>QuestionC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pearson Correlation</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>11.000</td>
</tr>
<tr>
<td>QuestionC</td>
<td>Pearson Correlation</td>
<td>.684*</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.020</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>11</td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.05 level (2-tailed).

Figure No. (5.6) Correlations
### Ratio Statistics

#### Case Processing Summary

<table>
<thead>
<tr>
<th>QuestionC</th>
<th>Count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
<td>45.0%</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>9.1%</td>
</tr>
<tr>
<td>38</td>
<td>1</td>
<td>9.1%</td>
</tr>
<tr>
<td>47</td>
<td>1</td>
<td>9.1%</td>
</tr>
<tr>
<td>55</td>
<td>1</td>
<td>9.1%</td>
</tr>
<tr>
<td>70</td>
<td>1</td>
<td>9.1%</td>
</tr>
<tr>
<td>130</td>
<td>1</td>
<td>9.1%</td>
</tr>
<tr>
<td>Overall</td>
<td>11</td>
<td>100.0%</td>
</tr>
<tr>
<td>Excluded</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>

#### Ratio Statistics for QuestionB / QuestionC

<table>
<thead>
<tr>
<th>Group</th>
<th>Price Related Differential</th>
<th>Coefficient of Variation</th>
<th>Median Centered</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.000</td>
<td>0.000</td>
<td>%</td>
</tr>
<tr>
<td>23</td>
<td>1.000</td>
<td>0.000</td>
<td>%</td>
</tr>
<tr>
<td>38</td>
<td>1.000</td>
<td>0.000</td>
<td>%</td>
</tr>
<tr>
<td>47</td>
<td>1.000</td>
<td>0.000</td>
<td>%</td>
</tr>
<tr>
<td>55</td>
<td>1.000</td>
<td>0.000</td>
<td>%</td>
</tr>
<tr>
<td>70</td>
<td>1.000</td>
<td>0.000</td>
<td>%</td>
</tr>
<tr>
<td>130</td>
<td>1.000</td>
<td>0.000</td>
<td>%</td>
</tr>
<tr>
<td>Overall</td>
<td>1.786</td>
<td>15.900</td>
<td>2973.1%</td>
</tr>
</tbody>
</table>

Figure No. (5.7) Ratio Statistics
Regression

Variables Entered/Removed

<table>
<thead>
<tr>
<th>Mode</th>
<th>Variables Entered</th>
<th>Variables Removed</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>QuestionD, QuestionB</td>
<td></td>
<td>Enter</td>
</tr>
</tbody>
</table>

a. All requested variables entered.
b. Dependent Variable: QuestionC

Model Summary

<table>
<thead>
<tr>
<th>Mode</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.474</td>
<td>.474</td>
<td>.343</td>
<td>33.45289</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), QuestionD, QuestionB
b. Dependent Variable: QuestionC

ANOVA

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Regression</td>
<td>2</td>
<td>4037.817</td>
<td>3.608</td>
<td>.078</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>8</td>
<td>1119.096</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>10</td>
<td>17226.200</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), QuestionD, QuestionB
b. Dependent Variable: QuestionC

t
<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>13.706</td>
<td>17.283</td>
<td>.769</td>
</tr>
<tr>
<td></td>
<td>Questions</td>
<td>1.227</td>
<td>1.236</td>
<td>.883</td>
</tr>
<tr>
<td></td>
<td>QuestionD</td>
<td>-1.411</td>
<td>.465</td>
<td>-1.078</td>
</tr>
</tbody>
</table>

a. Dependent Variable: QuestionC

Figure No. (5.8) Regression
**RELIABILITY**

/VARIABLES=QuestionA QuestionC

/SCALE(*ALL VARIABLES*) ALL

/MODEL=ALPHA.

**Reliability**

[DataSet 0]

**Scale: ALL VARIABLES**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>11</td>
<td>100.0</td>
</tr>
<tr>
<td>Valid</td>
<td>11</td>
<td>100.0</td>
</tr>
<tr>
<td>Excluded</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Case Processing Summary

a. Listwise deletion based on all variables in the procedure.

**Reliability Statistics**

<table>
<thead>
<tr>
<th>Cronbach's Alpha</th>
<th>N of items</th>
</tr>
</thead>
<tbody>
<tr>
<td>.811</td>
<td>2</td>
</tr>
</tbody>
</table>

Figure No. (5,9) Reliability
T-TEST
/PRETPVAL=0
/MISSING=ANALYSIS
/VARIABLES=QuestionB QuestionC
/CRITERIA=CI(.9500).

T-Test

| Dataset0 |

<table>
<thead>
<tr>
<th>One-Sample Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>One-Sample Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Value = 0</td>
</tr>
<tr>
<td>t</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>QuestionB</td>
</tr>
<tr>
<td>QuestionC</td>
</tr>
</tbody>
</table>

Figure No. (5.10) T-Test
APPENDIX (B)

TREATIES AND CONVENTIONS

Note

The following international conventions and treaties have been organised alphabetically by the hosting capital city of the country followed by the title and date.

Bangui Agreement on the Creation of an African Intellectual Property Organisation, March 2, 1977;

Banjul Protocol on Trademarks, March 6, 1997;

Brussels - Convention Relating to the Distribution of Programme-Carrying Signals Transmitted by Satellite, May, 1974;


Geneva - Convention for the Protection of Producers of Phonograms against Unauthorised Duplication of Their Phonograms, 1971; UNESCO.

Geneva - Universal Copyright Convention, with Appendix Declaration relating to Articles XVII and Resolution concerning Article XI, 6 September 1952; UNESCO.

Geneva - WIPO Internet Treaties, September 1999;

The Hague - The Hague Agreement Concerning the International Deposit of Industrial Designs of November 6, 1925;
Harare Protocol on patents and industrial designs in December 1982;

Lisbon - Agreement for the Protection of Appellations of Origin and their International Registration, as revised and amended, September 28, 1979;

Locarno - Agreement Establishing an International Classification for Industrial Designs, as amended on September 28, 1979;


Madrid - Madrid Agreement for the Repression of False or Deceptive Indications of Source on Goods, as of 1967;

Madrid - Agreement Concerning the International Registration of Marks of April 14, 1891, as revised and amended on September 28, 1979;

Madrid - Multilateral Convention for the Avoidance of Double Taxation of Copyright Royalties, with model bilateral agreement and additional Protocol. 13 December 1979; UNESCO.

Madrid - Protocol Relating to the Madrid Agreement Concerning the International Registration of Marks, adopted at Madrid on June 27, 1989;

Marrakesh Agreement Establishing the World Trade Organisation, including the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), of April 15, 1994;
Nice – Agreement Concerning the International Classification of Goods and Services for the Purposes of the Registration, as amended 1979;

Paris – Convention for the Protection of Industrial Property of March 20, 1883, as revised and amended, September 28, 1979;


PCT – Patent Cooperation Treaty, as in force from January 1, 2004;

PLT – (Patent Law Treaty)

Rome – International Convention for the Protection of Performers, Producers of Phonograms and Broadcasting Organisations, October 26, 1961;

Strasbourg – Agreement Concerning the International Patent Classification of March 24, 1971, as amended on September 28, 1979;

TLT-Trademark Law Treaty adopted at Geneva on October 27, 1994;


UPOV – International Convention for the Protection of New Varieties of Plants as last revised on March 19, 1991;

Vienna – Vienna Agreement Establishing an International Classification of the Figurative Elements of Marks, as amended on October 1, 1985;


WCT - WIPO Copyright Treaty and Agreed Statements Concerning the WIPO Copyright Treaty December 20, 1996;

APPENDIX (C)

CHRONOLOGY OF LIBRARY TECHNOLOGY


1919  American Library Association (ALA) adopts its first resource sharing code.

1952  ALA revises interlibrary loan code, adopting standardised interlibrary loan form.

1969  The first online public access catalogue is in use at the IBM Advanced System Development Division library.

1971  Libraries start sharing cataloguing resources electronically.

1979  Libraries begin lending and borrowing resources using a computer-driven interlibrary loan system.

1992  Librarian Jean Armour Polly coins the phrase ‘surfing the Internet’.

1994  Library Web sites launch at Virginia Tech University, the University of Michigan and the US Naval Research library.

1995  Jenny Levine creates the first library technology blog.
1996 The Internet Archive has been launched in San Francisco. The National Library of Australia began Web archiving.

1998 Bill Drew at the State University of New York at Morrisville offers real-time reference service using instant messaging (IM).

2000 OCLC introduced the ILLiad software for managing integrated interlibrary loan services (OCLC, 2000, op. cit.).

2006 Libraries start providing services in Second Life, an online, 3D virtual world.

2006 Launch of WorldCat.org—sharing the library holdings of more than 10,000 libraries on the Web.

2007 More than 25,000 videos on YouTube tagged or described with the term ‘library’ or ‘librarian’ as of September 2007 (OCLC, 2007, op. cit.).
## APPENDIX (D)
### DIGITAL STORAGE UNITS

<table>
<thead>
<tr>
<th>Unit</th>
<th>Abbr.</th>
<th>Size in Unit</th>
<th>Size in Bytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kilobyte</td>
<td>KB</td>
<td>1024 B</td>
<td>1,024</td>
</tr>
<tr>
<td>Megabyte</td>
<td>MB</td>
<td>1024 KB</td>
<td>1,048,576</td>
</tr>
<tr>
<td>Gigabyte</td>
<td>GB</td>
<td>1024 MB</td>
<td>1,073,741,824</td>
</tr>
<tr>
<td>Terabyte</td>
<td>TB</td>
<td>1024 GB</td>
<td>1,099,511,627,776</td>
</tr>
<tr>
<td>Petabyte</td>
<td>PB</td>
<td>1024 TB</td>
<td>1,125,899,906,842,624</td>
</tr>
<tr>
<td>Exabyte</td>
<td>EB</td>
<td>1024 PB</td>
<td>1,024,152,921,504,607,870,976</td>
</tr>
<tr>
<td>Zettabyte</td>
<td>ZB</td>
<td>1024 EB</td>
<td>1,024,180,591,620,718,458,879,424</td>
</tr>
<tr>
<td>Yottabyte</td>
<td>YB</td>
<td>1024 ZB</td>
<td>1,024,208,925,819,615,701,892,530,176</td>
</tr>
</tbody>
</table>

Source:
Appendix (E)

Letters of Permissions for Quotation

Dear Sir,

I would like to confirm that I have given permission to Mr. Gafar Ibrahim to quote some information from my thesis "A Records Management Programme for the Sudan National Records Office (SNRO) submitted as partial fulfillment for the award of Master Degree to University College London (UCL), University of London (UL), School of Library, Archive and Information Studies.

Thank you for your kind consideration

Should you require any further information please do not hesitate and contact me

Wafa Mohamed Osman
Documentalist and Information Manager
Commission for World Trade Organization Affairs
Director of Library and Documentation Center

WM Osman

--- On Mon, 7/21/08, Gafar.Ibrahim@ghd.com.au <Gafar.Ibrahim@ghd.com.au> wrote:
From: Gafar.Ibrahim@ghd.com.au <Gafar.Ibrahim@ghd.com.au>
Subject: Re: Letter of Permission for Quotation
To: wafamohkh@yahoo.com
Date: Monday, July 21, 2008, 12:47 AM

(See attached file: Permission for Quotation.doc)
Gafar Ibrahim
Librarian & Information Officer/Translator
Dear Gafar Ibrahim,

You are very welcome to make quotes from my dissertation.

Good luck with your research proposal!

Best regards,

Lennart Björneborn
Associate Professor, PhD
Information Interaction & Information Architecture
Royal School of Library and Information Science
Birketingsgade 6, DK-2300 Copenhagen S, Denmark
http://www.idb.dk/ib/

-----Oprindelig moddelelse-----
Fra: Gafar.Ibrahim@ghd.com.au [mailto:Gafar.Ibrahim@ghd.com.au]
Sendt: 14. juli 2008 08:46
Til: Lennart Björneborn
Emne: Permission for quotation

Dear Dr Lennart Björneborn,

Glad to write to and happy to read your valuable contribution to the webometric field. Actually, I need your permission to quote some facts from your dissertation posted on the web for developing a research proposal in Web archiving. As you know documentation of dissertations is essential if permission from the author and full citation is provided.

Looking forward to hear from your response.

Regards,

Gafar Ibrahim
Librarian & Information Officer/Translator

GHD | CLIENTS | PEOPLE | PERFORMANCE

T (+974) 4289541 | M (+974) 560 32 64 | F (+974) 444 6127 |
Gafar.Ibrahim@ghd.com.au
38 Dhow Street Al Sadd City Plaza Building P.O. Box 14352 Doha Qatar |
www.ghd.com.au

GHD serves the global markets of: Infrastructure | Mining & Industry |