INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES

Vol 7, Issue 11, (2017) E-ISSN: 2222-6990

# Development of a Generic Model for the Preservation of Primary Research Data Based Digital Resources Life Cycle

Irwan Kamaruddin Abdul Kadir and Alwi Mohd Yunus

Faculty of Information Management, Universiti Teknologi MARA

UiTM Selangor, Malaysia

**DOI Link:** http://dx.doi.org/10.6007/IJARBSS/v7-i11/3432

***Published Date:*** 11 November 2017

**Abstract**

This paper discusses the development of a generic preservation model for the preservation of social science primary research data and records created by research institutions and related organizations in Malaysia. This paper also presents the analysing of various best practice pragmatic examples policies, guidelines and models on recordkeeping functional requirements and data archiving. The result of the analysis leads to the analysis of the InterPARES Chain of Preservation (CoP) model and United Kingdom Data Archives (UKDA) Life Cycle model that underpin the construction of the generic model.

**Keywords:** Preservation Model, Research Data, Ukda, Life Cycle

## 1. Introduction

In order to set the context of the study, this Chapter begins with an analysis of the records life cycle and continuum concepts as every records including that of primary data and records for research should be managed in accordance to these concepts. The basis of the continuum concept is the functional requirements of recordkeeping systems with the ability to manage born digital records from creation through their preservation. In order to identify international best practices guidelines, policies and models on functional requirements of electronic records management, the next section of this Chapter presents an analysis of international projects devoted to the preservation of born digital records inclusive of primary research data and records of research in social science. Six international projects specifically on recordkeeping functional requirements were identified and analysed.

The following section of this paper discusses the reasons for adopting the CoP and the UKDA Life Cycle models and subsequently followed by the process of developing the generic preservation framework for the preservation of social science primary research data created by research institutions and related organizations in Malaysia. This is to fulfil the first aim of the study: To examine and analyse best practice models for preservation of digital primary research data in social science research through the process of identifying suitable models developed for the preservation of primary research data; examining and analysing best

practice models for the preservation of primary research data; and developing a generic preservation framework to be used as a framework for the construction of a model suitable for the Malaysian case.

The records continuum concept was developed by the Australian records and archival community. Before the emergence of the continuum concept, records practitioners the world over have conveniently used the records life cycle concept and it was regarded as one of the core concepts in records management. It is widely accepted by records management professionals as an important concepts to underpin total records management program best practice, so much so Pedley (1992:11, 64) equated the term with theory. The life cycle concept was developed in the USA by records practitioners in the 1930s. Through the life cycle concept, every record passes through various stages in its lives, be it paper or electronic. Every record has to be created, used, maintained and stored, appraised, and finally either be destroyed or preserved in an archive (Smith, 1996; Robek, Brown & Stephens, 1996; Gill, 1993; Schwartz & Hernon, 1993; Hare & McLeod, 1997; Shepherd & Yeo, 2003).

The life cycle concept has long been established and fits well with paper-based records management. Questions regarding the suitability of the life cycle concept to be used in managing born digital or electronic records have been much debated by the various experts, records practitioners and academicians (Irwan Kamaruddin, 2014). As a result of the debate, the Australian records management community came with an idea of a continuum concept and until the time of writing, the literature suggests that it seems to be the best concept for the management of born digital records (Duff, 1995; Upward, 1996; McKemmish, 1997; Marshall, 2000; Mutero, 2011). The aim of this paper is to review current development of a generic model for preservation of primary research data based life cycle of digital resources. In order to achieve this aim, we review digital preservation project at international level which in line with digital resources life cycle.

## 2. International Projects Devoted To Digital Preservation

The projects illustrated in Table 1 addressed digital preservation in various domains such as 1) e-government, scientific and research data, geospatial information, electronic health records, audio and video, media archives, government and cultural archives, corporate data archives; 2) domain specific adaptation of Open Archival Information System (OAIS) Reference Model; 3) digital disasters and recovery; 4) digital forensic for authenticity of information; 5) technological infrastructure development for trusted digital repositories; 6) trusted digital repositories in cloud; 7) digital preservation for legal compliances; 8) the challenges and techniques for maintaining trust in electronic records over long period; 9) integrating e-record management with e-government; 10) approaches for long term sustainability of digital repositories; 11) data mining and repurposing of digital information stored in digital repositories; 12) long term storage solutions with techniques for optimization and efficacy; 13) preparing for audit and certification of digital repositories; 14) design of domain specific digital preservation policy framework; 15) curriculum design for digital preservation.

One of the more theoretical projects that analysed a couple of components of digital records was the Investigating Significant Properties of Electronic Content(InSPECT) project. Supported by the Arts and Humanities Data Service (AHDS), the project was scheduled to operate from April 2007 through September 2008 and aims to “examine the whole concept of significant properties, determine which properties are significant for a range of object types and assess the importance of each of these for future representation of the object…”. Despite

the dearth of current, on-going projects solely devoted to digital records, a more recent trend has been to take a more granular approach to the issue. The SHERPA DP2 Project and the LIFE Project (Davies, Ayris, McLeod, Shenton &Wheatley,2007) can serve as examples in the area of institutional repositories while in the area of electronic records management, the InterPARESCoP Model 2 provides a comprehensive and well-documented process. More recently, projects such as the "Digitale Archivering in Vlaamse Instellingen en Diensten" (DAVID) and "Managing the Digital University Desktop" (MDUD) along with several individual studies have further explored this issue (Pennock, 2006). While the MDUD project has only made available a couple of presentations and an online tutorial regarding the management of e-mail, the DAVID project has elaborated on at least one issue that is at the core of InterPARES: authenticity (Force, 2009). In his report "Filing and Archiving E-mail," Boudrez (2006) asserted that, to ensure the authenticity of e-mails, "all essential components [must] … be archived," including their archival context, their content (subject and attachments) and their structure (header and message field). In his assessment, Boudrez offers a preservation solution based on a combination of Extensible Markup Language (XML) and migration—a process currently being used by the Antwerp city archives. These projects aside, it may be only assumed that, with the increasing reliance on e-mail, additional projects will emerge and continue to address issues and concerns related to this type of digital record (Force, 2009).

One of the dominant current trends among digital preservation projects is to focus on developing information systems that ingest, manage and retrieve "data," "digital objects," "digital information" and/or "digital content" based on the Open Archival Information System (OAIS) model. Where InterPARES 2 based its Chain of Preservation Model on the OAIS model, adapting it to archival theory and methods, 35 other projects have followed the model more literally, incorporating also its language. This has resulted in an often confusing lexicon of digital preservation terminology. For example, the Co-operative Development of a Long-Term Digital Information Archives (KOPAL), which is currently developing tools to ensure the long-term access to electronic publications, uses the terms "digital documents" and "digital data" interchangeably in its mission statement, while the "Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval" (CASPAR) project explores ways to handle the preservation of the digital resources of many user communities (Factor et. al, 2007).The "Securing a Hybrid Environment for Research Preservation and Access Development Partner" (SHERPA DP2) project, which is devoted to preserving electronic publications, works to develop its "OAIS-based distributed preservation model … to interact with repositories holding different and varied types of digital content." (Force, 2009).

Table 1.
*Preservation Issues Addressed by the Digital Preservation Projects.*

| **Research Projects** | **Year of Commencement** | **Preservation Issues Addressed** |
|---|---|---|
| Victorian Electronic Records Strategy (VERS) | 1990s | System Emulation |
| CURL Exemplars in Digital Archives (CEDARS) | 1990s | Preservation of bit stream of digital object in different format |
| Creative Archiving at Michigan and Leeds: emulating the old on the new (CAMiLEON) | 1990s | Emulating the old system on the new system |

| | | |
|---|---|---|
| University of Pittsburgh project | 1990s | Functional requirements for evidence in recordkeeping |
| University of Indiana project | 1990s | Testing the Pittsburgh functional requirements |
| University of Yale project | 1990s | Testing the Pittsburgh functional requirements |
| National Archives of Australia preservation of government records in electronic environment | 1995 | Statement of common position on electronic recordkeeping |
| US Department of Defence (DoD) | 1997 | Functional requirements for creating, handling and preserving electronic records |
| UKDA projects | 1998 | System for the preservation and open access of primary research data and records in social science disciplines |
| PRO'S Electronic Records Management Systems | 1999 | Functional requirements for electronic records management systems: Part One – Statement of Requirements |
| NARA's electronic records project | 1999-2000 | Systems functions for electronic recordkeeping and electronic records management system – Fast track guidance |
| National Archives of Malaysia eSPARK project | 2003 – 2007 | Preparation of working tools such as guidelines, procedures for ERM in Malaysian government and AMS |
| Network Excellence on Digital Libraries (DELOS) | 2004-2007 | To integrate research activities and disseminate knowledge of digital library technologies |
| The Long-Term Preservation Metadata for Electronic Records (LMER) | 2005 | A metadata schema for long-term preservation of electronic resources |
| Minnesota Historical Society E-Legislature | 2005 – 2007 | Testing technologies available to preserve electronic records of the Minnesota legislature |
| Effective Strategic Model for the Preservation and Disposal of Institutional Digital Assets (ESPIDA) | 2005 – 2007 | Cost models for the preservation of digital assets include primary research data |
| DLM Forum of the European Commission | 2002 - 2007 | Model requirements for electronic recordkeeping (MoReq) |
| Clever Recordkeeping Metadata Project (CRKM), Monash University | 2007 | Recordkeeping metadata standards |

| | | |
|---|---|---|
| Life-Cycle Information for E-Literature (LIFE) | 2007 | The life cycle generic preservation model |
| Investigating Significant Properties of Electronic Content (InSPECT) Project | 2007-2008 | Theoretical project analysing the components of digital records |
| Digitale Archivering in Vlaamse Instellingen en Diensten(DAVID) | 2007/2008 | Preservation of authentic archival records |
| Managing the Digital University Desktop (MDUD) | 2007/2008 | Filing and archiving e-mail |
| Co-operative Development of a Long-term Digital Information Archives (KOPAL) | 2007/2008 | Tools to ensure the long-term access to electronic publications |
| Cultural, Artistic and Scientific Knowledge for Preservation, Access and Retrieval (CASPAR) | 2007/2008 | OAIS-based distributed preservation model, access and retrieval of cultural and scientific knowledge |
| Securing a Hybrid Environment for Research Preservation and Access Development Partner (SHERPA DP2) | 2007/2008 | The concepts of institutional repository |
| Romanian Social Data Archive (RODA) | 2007/2008 | Preservation of authentic digital records |
| Service-Oriented Architecture for Preservation and Ingest of Digital Objects (SOAPI) | 2007 – 2008 | An metadata architecture and toolkit for partially automating preservation and ingest workflows in digital repositories |
| Digital Repository Infrastructure Vision for European Research (DRIVER) | 2009 | Built testbed for knowledge infrastructure of the European Research Area for data storage and data transport |
| Preservation E-print Services (PRESERV) | 2009 | Methods to preserve electronic publications |
| Exploring Collaborations to Harness Objects in a Digital Environment for Preservation (ECHO Depository) | 2009 | Tools for the preservation of digital objects |
| Data Preservation Alliance for the Social Science (Data-PASS) | 2009 | To identify, acquire and preserve data at risk of being lost to the social science research community |
| Preservation and Long-term Access through Networked Services (PLANETS) | 2009 | Tools and services to ensure long-term access to organizations valued digital content |
| Metadata Encoding and Transmission Standard (METS) | 2009 | Standards on metadata encoding and transmission |
| Image Spatial Data Analysis Group (ISDA/Ip2Learn) | 2009 | Tools and systems for the preservation of spatial data |

| | | |
|---|---|---|
| International Research Project on Permanent Authentic Records in Electronic Systems (InterPARES) | 1999 – 2012 | Chain of Preservation model for the preservation of the integrity and authenticity of electronic records through the life cycle and continuum concepts |
| LIBER Working Group on E-Science and Research Data Management | 2012 | To support research data management within the European E-Science Working Group |
| NDIIPP Preserving US State Government Information Projects | 2012 | Tools and systems for preservation of government records |
| DataUp Project | 2012 | Developing an open-source web application for creation of standardized metadata |
| Demystifying Born Digital Project | 2012 | Management of born digital materials |
| Web Archiving Project | 2012 | Tools for preservation of records on the web for archival purpose |
| Trust and Digital Records in an Increasingly Networked Society | 2013 | Concepts and tools for Preservation of records in the clouds |

Based on the careful analysis of the various products of the digital preservation projects as illustrated in Table 1, it was found that in order to preserve born digital records they must be captured in their entirety and appraised at their creating stage by using an effective tool that is a recordkeeping system with specific functional requirements. Functional requirements are usually very specific and measurable. In capturing the electronic record in its entirety, the issue of long-term preservation is actually tied into the creation and capturing of the electronic records as record managers must look to apply retention and disposition instructions from the earliest possible moment during the point of creation in a recordkeeping system (Walker, 2002; Duranti, 2013). The primary research data and records for research must be created in an electronic recordkeeping system with appropriate functional requirements so that they can be preserved and can be accessed when needed for reuse from the point of their creation.

*2.1 International Best Practices Guidelines, Policies And Models For Developing Functional Requirements Of Electronic Records Management*

The purpose of the analysis of national and international policies, guidelines and models best practices addressing the main functional requirements of ERMS systems aimed at confirming the functional requirements identified in the literature review showed in Table 1. The analysis identified common themes in the guidelines and resulted in a set of list of the main functional requirements of ERM. Essentially, the list provides a frame of reference against which to measure and construct the generic preservation framework for the preservation of primary data and records for research.

Firstly, the institutions and organizations listed are leading advocates for developing, establishing and providing authoritative guidance on all aspects of ERM and digital preservation. Secondly, the organisations are all leading advocates from different countries around the world, thereby offering a wider perspective of the functionality of ERM systems.

Thirdly, while the organizations listed produced many policies, guidelines and models regarding the management of born digital records, those identified addressed the whole functionality of ERM systems inclusive of the preservation aspects. The writer thought that it is important to include only policies and guidelines addressing all aspects of ERM in order to allow a more comparison between the policies, guidelines and models and their contents. Table 2 illustrates the main functional requirements of ERM identified in each document.

Table 2.
*The Main Functional Requirements of ERMS for Managing and Preserving Born Digital Records.*

| NFTG | DoD | NAACM | UKNA ERMS | EC MoReq | InterPARES CoP |
|---|---|---|---|---|---|
| Declare a record | Creating file plans | Create records | Records classification | Create classification scheme | Manage framework for CoP |
| Capture a record | Identifying records | Capture records | Declaration of records | Manage control and security | Select and preserve records |
| Maintain and use record | Filing records | Manage content of records | Capture records | Manage retention and disposal of records | Managing the creation of records |
| Record security | Storing records | Manage context of records | Capture metadata | Capturing records | Manage the receipt of records |
| Facilitate search and retrieval | Scheduling records | Manage structure of records | Manage control versions of records | Referencing /Identifier of records | Monitor performance of record making system |
| Manage access | Searching and Retrieving records | Maintain records in the long-term | Manage records retention | Searching of records | Facilitate access |
| Preserve records | Transfer records | Manage metadata of records | Manage disposition of records | Retrieval of records | Carry-out disposition |
| Create audit trail | Providing audit trails | Manage audit | Auditing and authentication | Rendering of records | Appraise records |
| Dispose records | Access control | Manage remote access | Manage import/export of records | Transfer of records | Acquire selected records |
| Transfer records | Destroying records | Manage findings aids for records | Manage migration of records | Destruction of records | Reproduce electronic records |

Having established the functional requirements of the ERM, the writer continues with the analysis of CoP and UKDA in order to justify why these two models are suitable for the construction of a generic preservation framework for the preservation of primary data and records for social science research. The CoP and UKDA were not part of Walker's (2002) model framework.

## 3. The Lifecycle Of Digital Resources

### *3.1 The Creation of Digital Resources of Research Projects Activities*

There are nine (9) data creation requirements namely: format, type, storage, medium, metadata, accessibility, security, program and policy/procedures. Data creation requirements are essential to ensure records created are being well-managed until its disposition. Born digital data and records of research activities will go through its phases of the lifecycle from creation, maintenance, appraisal, disposition, preservation and continuing access. If at the creation stage, the records creation requirements are not being fulfilled, the tendency of the records are lost and to be less authentic, reliable and usable is high. The purpose of having the nine data creation requirements at the early stage in records management (which is creation) is to allow better management at every phase of record-keeping processes.

Records format can be in various formats such as word processing documents, images, mathematical, graphics, presentation and other electronic data format. For instance, word processing document can be in Microsoft Office document (doc) format, Office Open XML document (docx) and latest in portable document format (pdf). Format of the records are essential to enable records to be retrievable and accessible. The data format must be supported by the data archives system of the organisation. Data type can either be alphabetic, numbers, alphanumeric and symbols. Combination of data types constructs the information stored in the database that is comprehensible by the users. Storage is one of the requirements in managing data archives as it becomes the "shelves" of the data kept. In electronic environment, database is the storage medium for the data to be kept.

Metadata for instance the title of the records, date created, date modified and other important information in records are essential to support the retrieval audit trail and the evidential of a record created, edited or deleted. Accessibility on the other hand, will enable the retrieval of information by the right person and at the right time. Program is software used to open the data stored in the data archive. For instance, document processing document is using Adobe Reader program that support the retrievable of information without allowing the information to be edited or deleted. Each record stored will have policies and acts to be followed to ensure it has the standardization in data storage and retrieval.

Each of the elements will be managed and used in every phase in the data lifecycle. The data lifecycle started from the creation of the records until the records are reuse as references by the same researcher or other researchers. At the creation of the records, elements of data format, type and program are managed together in the data archive. This is because those elements are needed even in the first phase in the data cycle to ensure the records are remained accessible until the records are being reuse. Usually, the data created will be used immediately right after it was been created. At this stage of usage, the elements of accessibility and security will play its roles.

### *3.2 Data and Collection Management and Preservation*

Data and collection management and preservation may involve a number of stakeholders who can fulfil different functions and roles. These functions and roles may be for

a fixed or indefinite duration and can involve direct or indirect participation in the process. Immediately after creation of the data and usually for a period after this the primary data creators and digitizers will be responsible for the management and short-term preservation of the resource. The resource can also be deposited or will be transferred at a subsequent point to institutions or internal departments which will support or assume responsibility for long-term preservation and access. These functions can be undertaken by internal departments within the digitizers where their organizations' roles extend to long-term preservation. Alternatively these functions will be achieved by offering to deposit with and/or acquisition of the resource by the institutional archives, copyright and deposit libraries, and academic archives.

In addition, digital information may be created as part of the process of collection building or collection management of a resource. This can be seen as an extension or supplement to data creation process and similar criteria will apply. Collections may be extended or new aggregations of resources created by licensing, copying or mirroring existing digital information created by others. New digital information can also be created in collection management processes e.g. the computerized cataloguing or digital research materials generated from existing resources in digital or traditional forms.

In some cases the resource or collections may be managed and preserved by administrative processes which we have described as "remote management". For dynamic constantly changing information, a single deposit and acquisition for long-term preservation may be inappropriate. In such cases digital information may remain with the data creator who will assume responsibility for updating and maintaining it. The primary data creator may be legally obliged or voluntarily abide by standards and procedures established by an external organization with established procedures for deposit. Decisions may be taken to periodically sample or copy the resource which will provide an archive of the resource at particular points in time.

"Active" resources which are still used by their creators in a current project or business process may be managed and preserved by a similar process of remote management in which the data creators abide by standards and procedures agreed with and monitored by an external organization. In such cases the data may be reviewed and selected for deposit and acquisition when it is no longer in an active phase of use by the data creator. Alternatively a copy of the data may have been deposited during this active phase but access may be denied or restricted for an agreed period.The organizations we have identified as "data banks", and to a more limited extent other organizational types, may also be involved as contractors in remote management of resources. They frequently manage resources under contract to others who retain legal responsibility for the resource and set terms and standards in the contract for their management.

### *3.2.1 Acquisition, Retention or Disposal*

Acquisition of a resource may involve decisions about collection policy, selection and rejection criteria, sampling methodology, collection levels, retention periods, disposal of part or all of a resource, selection for long-term preservation, and which data resources should be accessioned into (or excluded from) a permanent collection or handled by remote management of the resource. It will also involve data evaluation - a nuts and bolts assessment of those data resources which are potential acquisitions and will determine how (even whether), and at what cost a data resource may be included in a collection and its fitness for its intended purpose. This process will be critically dependent on or affected by decisions made

when the resources were created: the formats and structures used, data quality and consistency, the existence of metadata and documentation, or the rights accompanying the resource. Decisions taken when the resource is acquired will subsequently shape the collection and impinge directly upon how it is catalogued and documented, managed, made accessible to end users, and preserved. The selection process occurs primarily when the resource is acquired but can be an iterative process. Decisions not to retain a resource or to transfer it to another organization can occur after an agreed review period or as the collection policies of an organization and its peers evolve and change over time.

*3.2.2 Data Management*

A suite of related decisions about how data resources are handled and described once they are included in a collection. How data is managed will depend upon how it has been created or supplied (e.g. in what format, with what documentation, and under what terms and conditions). Data management options will accordingly be constrained by decisions taken when data is created or selected for inclusion in a collection and by the funding and technology available to the organization. They will also constrain data use and preservation options.

*3.2.3 Data Structure, Format, Compression, and Encoding*

How data is formatted (written to magnetic media), compressed, and encoded (i.e. how internal semantic or syntactic features are represented) will determine its portability across hardware and software platforms and how it may be stored, preserved, manipulated, and subsequently enriched.

*3.2.4 Data Description and Documentation*

The information supplied about a data resource's structure, contents, context, provenance, and history. The information will normally be in two parts; information which was created with the resource such as users' manuals and data dictionaries or provided to document its transfer; and secondly new digital information created when existing resources in traditional or digital form are catalogued or supplemented by research. It influences how a resource is located, managed, and used, and frequently reflects data acquisition decisions (notably as they reflect what documentation is supplied for a resource, how it is supplied, and who supplies it), and the subject or sectoral documentation standards and practices of the creators and curators of the resource. It will also be contingent upon the resources in terms of cataloguing staff and expertise available to the managing agency.

*3.2.5 Data Storage*

It involves organizational decisions about whether collections or parts of collections are stored centrally or distributed across several sites, contracted out to a data bank, or the technical decisions about what magnetic media and hardware platforms, physical security, refreshing or replacement of storage media, and contingency procedures, are used. Options are constrained by the resources' structure format, compression, and encoding; by whether the resource is dynamic or fixed in its nature; the need to maintain authenticity and integrity of the resource; and also upon the relative emphasis given to their use and/or preservation. Accordingly data storage decisions together with the available funding and technologies can constrain data creation or acquisition and help to determine how (even whether) and to what extent a data resource once included in a collection can be preserved and/or used.

*3.2.6 Data Preservation*

A suite of strategic and procedural decisions which together with other aspects of data management help to ensure that the content, context and authenticity of a data resource survives through time and changing technologies with minimal loss in its information content, functionality, and accessibility.

*3.2.7 Data Use*

Data use can occur immediately after its creation and for an indefinite period thereafter. Its use can be to fulfil its primary purpose when created, involve subsequent secondary analysis, or inclusion in a collection developed to fulfil other aims. The primary data creators, digitizers, funding agencies, publishers, institutional archives, copyright and deposit libraries, academic archives and their user communities may all be involved in data use or defining and servicing user requirements. Use of the data will be highly contingent on the decisions made and circumstances surrounding creation, management and preservation of the resource; the rights management and economic framework which applies, and the approaches taken to identify and reconcile the needs of different stakeholders.

How data is delivered to and used by end users will be contingent upon: how and why it was created or acquired; agreements to co-operate, share or exchange data between different institutions; conditions and procedures required to meet legal and economic requirements; how/where it is stored; and upon what software and hardware is needed to access it. Its use over extended periods of time will also be contingent on decisions made on data management and preservation.

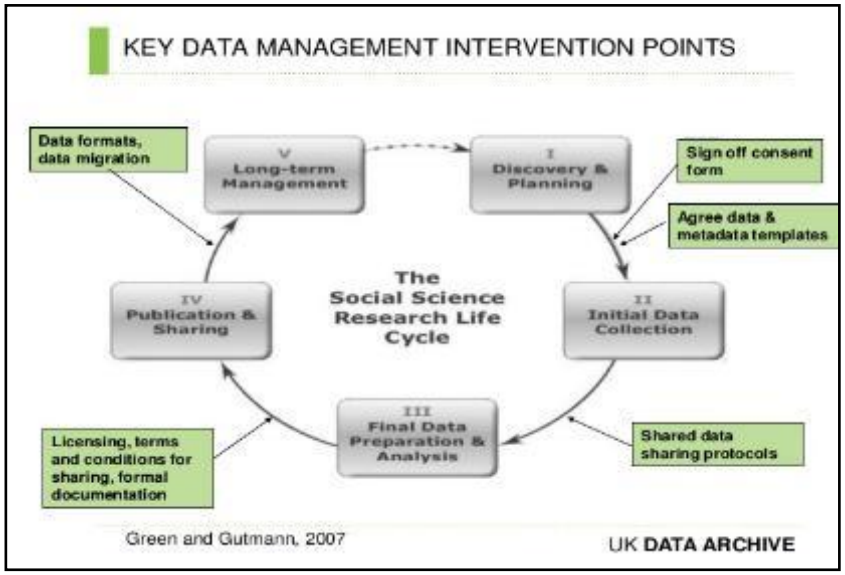


Fig. 1. The Social Science Research Life Cycle.

Table 3.
*Life Cycle of a Digital Resource of Research Project Activities.*

| | |
|---|---|
| **Data creation** | Decisions made when the digital resource was created-often outside the control of the collection manager, but having a major impact on the options subsequently available. |
| **Data management and preservation** | The processes by which management and preservation of data are made and conducted. |

| | |
|---|---|
| **Acquisition, retention or disposal** | Decision based on the digital resource's, compressed and encoded. |
| **Data structure** | How a digital resource is formatted, compressed and encoded. |
| **Data description and documentation** | The extent to which the digital resource's structure, content, provenance and history have been documented. |
| **Data storage** | The computer hardware and media used to store the digital resource. |
| **Data preservation** | Safeguarding the information content of any digital resource from the ravages of time, technological change and decaying magnetic media. Different strategies are appropriate for different data types and structures. Preservation requirements will impinge on how digital resources are structured, documented, stored and validated and possibly even on the conditions and methods by which digital resources can be accessed by end users. |
| **Data use** | Decisions on how digital resources are to be delivered and used; will be influenced by how they were created and will influence how they are managed. |
| **Rights management** | Intellectual property rights, data protection and confidentiality issues; need to develop both acquisition licences and distribution licences and implementation procedures. |

Table 4.
*UKDA Framework in Practice Compared to Data Bank Services.*

| **Stage in Life-Cycle; Data Bank** | **UKDA** |
|---|---|
| **Data Creation** | |
| Acting on a contract basis to manage data at the bit-stream level, with no interest in a data resource's future usage, and compelled for economic advantage to offer the same service to all, the data bank has little interest in how, why, or for whom deposited data are created. | This unique perspective is apparent in the core services offered at UKDA which accessions and stores data created in a variety of different standard and non-standard formats. Where UKDA's work with the UK National Archives is concerned, UK National Archives guidelines pertaining to the management of computer-readable datasets mitigate to a larger extent the need for that role being taken up by the data bank. |
| **Data Acquisition** | |
| The data bank operates on a cost for quantity economic model and so its role in data selection is limited. | UKDA departs some way from the norm in its work for the UK National Archives. Although UKDA must archive all data resources and information systems deposited by the National Archives, it does exercise some influence, in discussion with the UK National Archives about accessioning priorities and costs, and with officials in government departments who are responsible for identifying and preparing records for long-term retention. |

| | |
|---|---|
| **Structure and Storage** | |
| Data banks leave responsibility for how data are formatted, encoded and compressed with depositors, though may regulate how (e.g. on what media) deposited data may be transferred. They are therefore largely unconstrained in the data structures they can accommodate and will not normally need to restructure data unless they are contracted by the depositor to perform content migration or data distribution functions or to provide access services. | UKDA will undertake these additional functions when engaged (and funded) to do so either by the record generating project, or by the record generating project, or by the designated University authority which may take responsibility for the long-term preservation of certain data resources. Government departments take account of data resources' physical and technical characteristics when selecting data for deposit. UKDA will also restructure data deposited by the UK National Archives since it is engaged to migrate them through changing technical regimes and make them accessible to users. |
| **Data Description and Documentation** | |
| With the exception of essential administrative information which is supplied by the data bank to locate, name, and record other vital statistics about deposited data, data documentation is left entirely to the depositor. | Again, UKDA's role is exceptional where UK National Archives data are concerned, since the UK National Archives has contracted out to it some functions in standardising, and enriching documentation that is supplied by depositor. |
| **Data Preservation** | |
| Data banks migrate data files through storage media to ensure their readability, but content migration (ensuring that data can be meaningfully represented by and accessed from contemporary platforms) is the responsibility of the depositor. The data bank will rely upon extensive computing infrastructure which may include large-scale computer servers, robotic tape libraries, etc. Preservation is based around the management of archive copies of the deposited data resources; that is, copies which are independent of any online representation they may have. | A preservation scenario: Archive copies are stored on industry standard digital tape or other approved media as may arise, and there will be multiple copies of any single data file, some stored on and others stored off site, preferably in temperature controlled and fire-proof safes or rooms. Off-site copies should be a safe distance from on-site copies to ensure they are unaffected by any natural or man-made disaster affecting the on-site copies. Archives copies may be written with different software to protect data against corruption from malfunctioning or virus-or bug-ridden software, and may be made to comparable magnetic media purchased from different suppliers to guard against faults introduced by the media's suppliers into their products or into batches of their products. Data files stored as archive copies will be migrated periodically to new media with that migration taking place within a minimum time which reflects the media supplier's estimate for the media's viability under prevailing climate conditions. In addition, media will be checked periodically for their readability. Such checking may be conducted automatically by archive |

| | |
|---|---|
| | systems according to parameters set by system operators. The integrity of data files may also be checked using checksum and similar procedures which may be implemented automatically by the archive system according to parameters set by system operators. |
| **Data Use** | |
| Beyond ensuring that depositors can recall their data on readable media, the data bank is unconcerned with re-use. User support is oriented exclusively toward depositors (typically also the data's sole users) and may include documentation about the service on offer, how it works, and how access to it may be acquired. | UKDA's position is complicated by its having been contracted to the UK National Archives to distribute holdings in its Computer Readable Data Archives. In this respect, to adopt functions more typically associated with an institutional or academic data archive. User support services are also complicated by the data bank's involvement in providing third-party access to UK National Archives-deposited data. |
| **Rights Management** | |
| Since the data depositor tends to be the sole user of data which are stored in a data bank, rights management is not a central concern. | Depositors take full responsibility for data they deposit in the UKDA. |

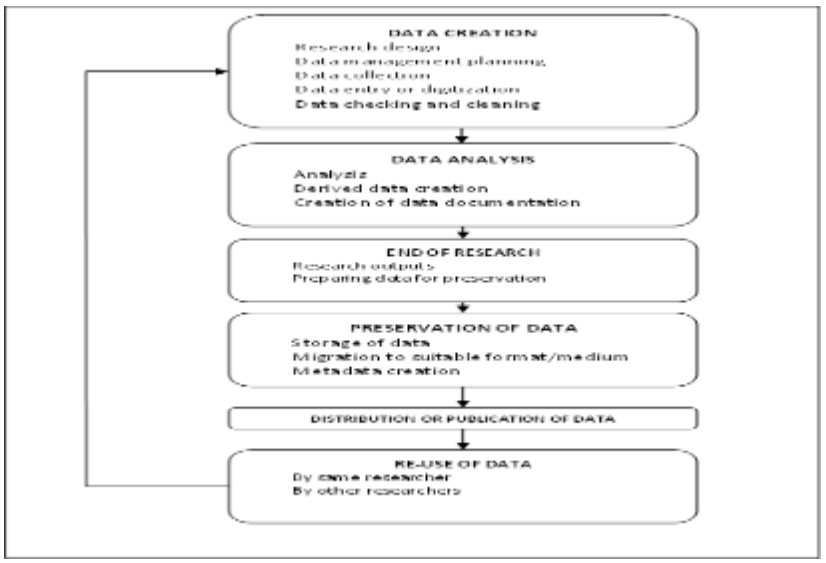


Fig. 2. UKDA Data Lifecycle Model.

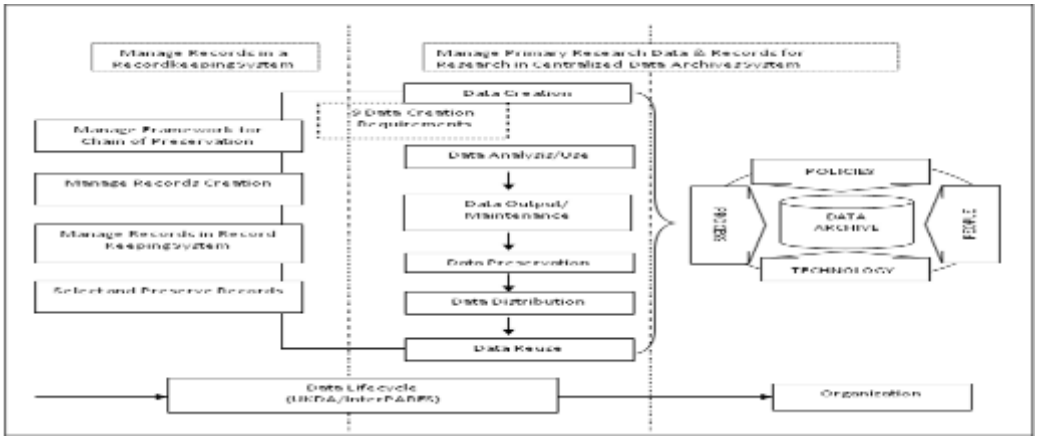

Fig. 3. First Level of Primary Research Data Preservation Framework.

The Chain of Preservation model (CoP) presents four main activities in managing records that are lengthily elaborated in previous pages in this chapter. Thus, for the creation of model that suit Malaysian cases, these four activities are implemented into the above diagram. The first activity is Managing Framework for Chain of Preservation. Secondly, Managing the Creation of Records followed by Manage Records in Recordkeeping System and lastly Select and Preserve Records. In creating data or records, there are some guidelines or requirements that are needed to be fulfilled to ensure the creation of the data are complete, comprehensive and accurate. The United Kingdom Data Archives (UKDA) presents nine data creation requirements that are needed to be included in the creation of data. These are format, type, storage, medium, metadata, accessibility, security, program and policy/procedures. Thus, in data creation phase, the nine requirements are mandatory. These nine data requirements crucially turn out to be necessitates in the creation of data for a superior management of the next phases of managing primary research data in centralized data archives system. By adopting the InterPARES recordkeeping functional requirements and UKDA data lifecycle, this generic preservation framework was developed in order to measure the existing practices of the Malaysian case. Different research organisation might have different data to be stored which depends on its functions and policy enforced. The organisation might also have different level of confidentiality of the data. However, the elements of the data will be the same as proposed in the generic preservation framework as well as the data lifecycle will also be the same as proposed by the InterPARES recordkeeping and UKDA models. In managing records in recordkeeping system, the elements and lifecycle phases closely related to one another for a creation, maintenance, usage, storage and preservation of an authentic, reliability and usability of the records.

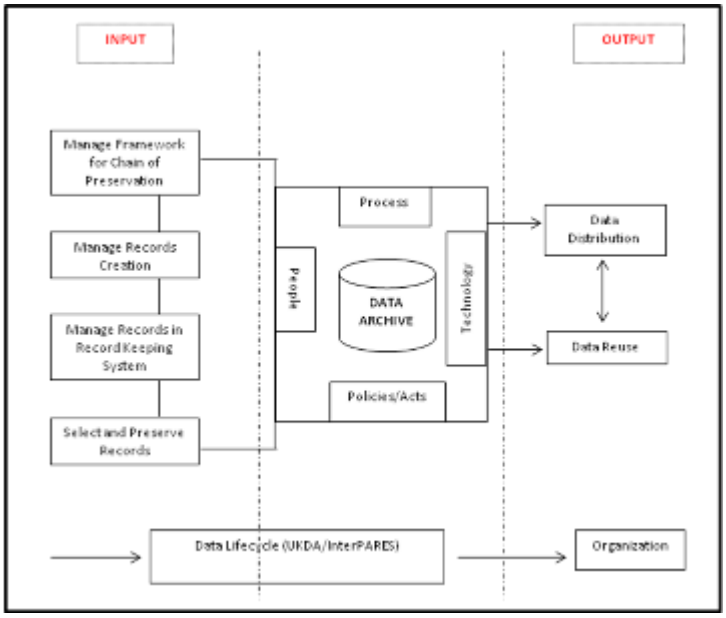


Fig. 4. Second Level of Primary Research Data Preservation Framework.

The evolution of the model showed that the four main activities become the core elements in this model. The four activities as suggested by the Chain of Preservation are needed to be managed and completed for accurate and reliable data and records are being stored in the data archive. As shown in the model, the four activities are following the records lifecycle as every record life is started with is creation, maintenance, appraisal and finally preservation. Thus, these records keeping system are being carefully managed and kept in the data archives with the existence of people to manage the records such as records manager, IT officers, programmer, research administrators and etc. Technology on the other hand includes computer hardware and software, database and server. Policies to regulate and administer the records stored and process to handle the process of data keeping.As data and record are systematically select and kept, the output of the whole processes are to distribute the records to the end-users or researchers. As records are being distributed to the users, it also being reuse by variety of users such as the creators of data and records, government agencies over time as long as the records are available and retrievable in the data archive.

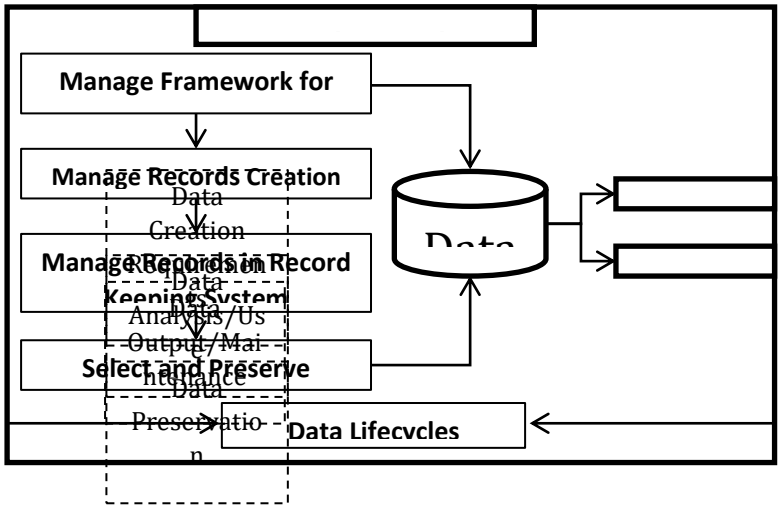


Fig. 5. Third Level of the Construction of the Generic Model for the Preservation of Primary Research Data and Records.

This diagram provides an extension of the previous model. The four core activities are comprise their own sub-activities before the records being kept in the data archives as permanent research data. In the activity of managing records creation, the nine data creation requirements are needed. The requirements of format, type, storage, medium, metadata, accessibility, security, program and policy/procedures are required to ensure a comprehensive creation of records are being done. The next activity as proposed by the Chain of Preservation is the Manage Records in Record Keeping System consists of two sub-activities which are data analysis and data output/maintenance. The fourth activity consists of the sub-activity of data preservation. Thus, from the previous diagrams, the separated four core activities and the cycle of the data management are being integrated into each appropriate core activities. For the data archive, the policy/procedures are being eliminated from the requirements as the policy and procedures are already highlighted during the early stages in the record keeping management which in the nine requirements of data creation. Thus, data archives only need three elements for proper management which are the people, process as well as technology. Finally, the data stored and kept will be used by the end-users or researchers and will re-use by the researchers as long as the data in the data archive remain accessible and retrievable.

Ultimately, the models are being designed according the UKDA as well as the CoP requirements that are globally accepted in the world of records management. The lifecycle of records and the continuum of records are also highlighted along the model from the creation until the final disposition takes place. In digital records management, the continuum of records is essential as the processes of the records are continuous from the keepers of the records until the phases of the records kept.

## 4. Conclusion

The continuum concept of recordkeeping with specific functional requirements identified through an analysis of best practice pragmatic examples provides the core elements of the management of born digital records. This is deemed necessary to capture the primary data and records for research in their entirety by using an ERM system with recordkeeping functional requirements in the creating agencies. The UKDA concepts on the other hand provides a continuum model of core functional elements of pragmatic examples which enable the preservation of these records so that they can be openly accessed for reuse by the public at large. The electronic recordkeeping functional requirements and the UKDA concepts underpinned the construction of the generic preservation framework for the preservation of primary data and records for research in social science.It is clear indeed that in order to preserve primary data and records for research there are a number of generic requirements or functions that can be grouped into a set of component elements and their supporting elements. Figure 4.10 on page 143illustrates those generic required elements that together form a coherent whole of the preservation of primary data and records for research in social science. This generic preservation framework will be used to measure the existing practices of the management and preservation of primary research data and records in social science in the surveyed Malaysian research institutions and related organizations. The next Chapter presents the analysis of quantitative data gathered in the first part of Phase 2 of the study aiming at identifying the existing practices of the management and preservation of the primary data and records in social science research of the Malaysian case.

**References**

Alemneh, D.G. & Hartsock, R. (2014). Theses and dissertations from print to ETD: The nuances of preserving and accessing those in music. In J.M. Krueger (Ed.). *Electronic Records and Resource Management Implementation in Diverse Environments* (pp 41-60). Hershey PA: IGI.

Aliza Ismail & Zuraidah Abd Manaf (2010). Malaysian cultural heritage at risk?: A case study of digitisation projects. *Library Review,59*(2), 107 – 116. Retrieved December 11, 2014,from http://dx.doi.org/10.1108/00242531011023862

Asma' Mohktar (2008). Polisi pengurusan rekod elektronik dalam sektor awam di Malaysia. *Jurnal Teknologi Maklumat & Multimedia **7**(2008),* 51 – 67.

Asmadi Mohammed Ghazali (2011). *Development of competency profile for records managers in the Malaysian federal ministries* (Doctoral dissertation). Universiti Teknologi MARA.

Arvidson, A. & Lettenström, F. (1998)**.** The kulturarw project-the swedish royal web archive. *Electronic Library, 16*(2), 105–108.

Aronson, B. & Glover, S. (2005). Journal access initiatives: Where are they? Why the World health organization? and What is HINARI? *Serials Review*, *31*(4), 279-283.

Azman Mohd Isa (2010). *Strategic management of records and risks for the sustainability of organizations* (Unpublished Doctoral dissertation). University of Glasgow,United Kingdom.

Ball, A., Day, M. & Patel, M. (2008). The fifth international conference on preservation of digital objects (iPRES 2008). *International Journal of Digital Curation*, *3*(2), 89-102.

Beagrie N. & Greenstein, D. (1998). A strategic policy framework for creating and preserving digital collections: A report to the Digital Archiving Working Group. Arts and Humanities Data Service. *Internet Scout Review*, *5*(2).

Becker, C., Kulovits, H., Guttenbrunner, M., Strodl, S., Rauber, A. & Hofman, H. (2009). Systematic planning for digital preservation: evaluating potential strategies and building preservation plans. *International Journal on Digital Libraries*, *10*(4), 133-157.

Bennett, J. C. (1997). A framework of data types and formats, and issues affecting the long term preservation of digital material. *British Library Research and Innovation* Report 50. London: British Library.

Catani, S. (2014). The preservation of electronic records: Decision acts signed by the department head. In J.M. Krueger (Ed.). *Electronic Records and Resource Management Implementation in Diverse Environments* (pp 318-333). Hershey PA: IGI.

Chen-Gaffey, A. (2014). Managing vendor records for monographic e-collection at a medium-sized academic library. In J.M. Krueger (Ed.). *Electronic Records and Resource Management Implementation in Diverse Environments* (pp 22-40). Hershey PA: IGI.

Chen, K. H. & Hsiang, J. (2009). The unique approach to institutional repository: practice of National Taiwan University*. Electronic Library, 27*(2), 204-221.

Cissne, K. (2014). NARA: A digitization case study. In J.M. Krueger (Ed.). *Electronic Records and Resource Management Implementation in Diverse Environments* (pp 306-317). Hershey PA: IGI.

Covey, D. T. (2010). Usage and usability assessment: Library practices and concerns, usage studies of electronic resources.*Council on Library and Information Resources.* Retrieved on January 20, 2012, from http://www.clir.org/pubs/reports/pub105/contents.html

Currall, J., Johnson, C. & McKinney, P. (2005). The organ grinder and the monkey: Making a business case for sustainable digital preservation.*EU DLM Forum Conference*, Budapest, Hungary.

Day, R. (2011).*How to write and publish a scientific paper*. (7thed.). Oxford: Greenwood.

Digital Curation Centre (UK) & Digital Preservation Europe (DPE). (2007). *Digital preservation, archival science and methodological foundations for digital libraries.* Retrieved June 1, 201, from http://www.gallbladder-research.org/media/media_113621_en.pdf

Duranti, L. (2011). Continuity and transformation in the role of the archivist. *Paper presented atInterPARES International Symposium*. 15 – 18 June, Kuching Sarawak, Malaysia.

Duranti, L. (2013). Trust in digital records: An increasingly cloudy legal area. *Computer Law & Security Review* 10/2013, *28*(5), 522-531.

Erwin, T. & Singer, J.S. (2009). The natural geospatial digital archives: A collaborative project to archive geospatial data. *Journal of Map & Geography Libraries: Advances in Geospatial Information, Collection & Archives,* 6(1),6-25.

Fresco, M. (1996). Long term preservation of electronic materials.*Report of a JISC/British Library Workshop as part of the electronic libraries programme (eLib)*, organized by UKOLN, 27-28 November 1995, University of Warwick. British Library R&D Report 6328. London: British Library.

Guercio, M. (2009). Keeping and preserving email. *Proceedings of InterPARES 3 Symposium*. 4-6 June, pp 143-195. Seoul: National Archives of Korea.

Gustavsen, I. (2009). The challenges of migration as a long-term preservation strategy: the findings of TEAM Norway and LongRec. *Proceedings ofInterPARES 3 Symposium*. 4-6 June, pp 279-300. Seoul: National Archives of Korea.

Haynes, D., Streatfield, D., Jowett, T. & Blake, M. (1997). *Responsibility for digital archiving and long term access to digital data*. British Library Research and Innovation Report 67. London: The British Library.

Hendley, T. (1998).*Comparison of methods and costs of digital preservation*. British Library Research and Innovation Report 106. London: The British Library.

Higgins, S. (2011). Digital curation: The emergence of a new discipline. *International Journal of Digital Curation*, *6*(2), 78-88.

Holdsworth, D. (2006). Strategies for digital preservation. *Digital preservation*, *64*(4), 32.

Irwan Kamaruddin Abdul Kadir (2014). *Electronic records management (ERM) practices at the Malaysian federal ministries: Exploring the fundamental organizational and professional issues* (Doctoral dissertation). UiTM.

Knight, S. (2010). Early learnings from the national library of New Zealand's national digital heritage archive project. *Program: electronic library and information systems*, *44*(2), 85-97.

Kruse, F. & Thestrup, J. B. (2013). Research libraries' new role in research data management, current trends and visions in Denmark. *LIBER Quarterly*, *22*.

Krueger, J. (2014). Building a digital repository for accreditation through courseware. In J.M. Krueger (Ed.).*Electronic Records and Resource Management Implementation in Diverse Environments* (pp 150-163). Hershey PA: IGI

Lauridsen, L. (2014). Topaz Japanese-American relocation center digital collection: A case study. In J.M. Krueger (Ed.).*Electronic Records and Resource Management Implementation in Diverse Environments* (pp 117-129). Hershey PA: IGI.

Lee, C. A., Kirschenbaum, M., Chassanoff, A., Olsen, P. & Woods, K. (2012). BitCurator: Tools and techniques for digital forensics in collecting institutions.*D-Lib Magazine*, *18*(5/6), 14-21.

McKemmish, S. & Gilliland, A. (2013). Archival and recordkeeping research-past, present and future. *Research methods: information systems, and contexts*, 79-112.

McLeod, J. & Child, S. (2003). *Managing primary research data and records for research in HE institutions* (Unpublished report). United Kingdom: Northumbria University.

McLeod, J., Hare, C.& Rusnah Johare (2004). Education and training for records management in the electronic environment- the (re)search for an appropriate model*. Information Research, 9*(3), 179.

McPeck, T.L. (2014). Electronic resources and next generation public library catalogs. In In J.M. Krueger (Ed.).*Electronic Records and Resource Management Implementation in Diverse Environments* (pp 1-21). Hershey PA: IGI

Murakas, R. & Rämmer, A. (2010). Social science data archiving and needs of the public sector: The case of Estonia. Estonian Social Science Data Archive (ESSDA).

Nurussobah Hussin (2013). *Functional requirements for the management of electronic court records in the Superior Court of Malaysia* (Doctoral dissertation). UiTM.

Pearce-Moses, R.& Davis, S. E. (2008). New skills for a digital era. In *Proceedings of a Colloquium sponsored by the National Archives and Records Administration, Society of American Archivists and the Arizona State Library, Archives and Public Records* (Vol. 31).

Ras, M. (2009). The KB e-depot: Building and managing a safe place for e-journals. *Liber Quarterly*, *19*(1), 44-53.

Rodrigues, P. & Pedron, C. D. (2013). Knowledge management on the implementation of a CRM project in a subsidiary company: A case study.*10th International Conference on Information Systems and Technology Management – CONTECSI.* 12 – 14 June 2013. São Paulo, Brazil.

Ross, S. (2012). Digital preservation, archival science and methodological foundations for digital libraries. *New Review of Information Networking*, *17*(1), 43-68.

Ross, S.& Gow, A. (1998). *Post-hoc rescue of digital materials. Digital archaeology? Rescuing neglected or damaged digital resources*. British Library Research and Innovation Report 108. London: The British Library.

Ross, S. & Hedstrom, M. (2005). Preservation research and sustainable digital libraries. *International Journal of Digital Libraries*, *5*(4), 317-325.

Royal Library of Sweden (2005). *Swedish websites.* Retrieved June 20, 2013, from http://www.kb.se/english/find/internet/websites/

Rusnah Johare, Alwi Mohd Yunus, Irwan Kamaruddin Abdul Kadir & Haslinda Mohamed. (2009). Managing primary research data and records for research in research institution and related organizations: Examples from the team Malaysia case studies. *Proceedings of theInterPARES 3 International Symposium*. Seoul, South Korea.

Rusnah Johare (2011). Case study 03-A study on digital preservation policy of PUSTAKA Negeri Sarawak. *Paper presented atInterPARES 3 Summit & Symposium*. 16-18 June. Kuching,Sarawak.

Sam, G. O. (2009). Toward developing a standard migration procedure. *Proceedings of InterPARES 3 Symposium*, 4-6 June, pp 19-54. Seoul: National Archives of Korea.

Shu-Fen, H.L. & Hsueh-Hua, C. (2014). Construction of a culture-rich database system for indigenous documentary records: Conceptual model, flexible classification, and

methodology. In J.M. Krueger (Ed.).*Electronic Records and Resource Management Implementation in Diverse Environments* (pp 61-93). Hershey PA: IGI.

Smith, M. (2005). Exploring variety in digital collections and the implications for digital preservation. *Library Trends*, *54*(1), 6-15.

Snell, C. (2014). ERMS druthers. In J.M. Krueger (Ed.).*Electronic Records and Resource Management Implementation in Diverse Environments*(pp 181-192). Hershey PA: IGI.

Stafford, D. & Flatley, R. (2014). Choosing and implementing an open source ERMS. In J.M. Krueger (Ed.).*Electronic Records and Resource Management Implementation in Diverse Environments* (pp 164-180). Hershey PA: IGI.

Stieglitz, S. (2014). The American memory project. In J.M. Krueger (Ed.), *Electronic Records and Resource Management Implementation in Diverse Environments*(pp 106-116). Hershey PA: IGI.

Utulu, S.C.A. & Akadri, A.A. (2014). A case of Redeemer's University adoption of instructional repository using the principles of electronic information management systems. In J.M. Krueger (Ed.).*Electronic Records and Resource Management Implementation in Diverse Environments (*pp 130-149). Hershey PA: IGI.

Voutssas, J. (2012). Long-term digital information preservation: Challenges in Latin America. *Aslib Proceedings, 64*(1), 83 – 96.

Wang, J. (2009). Challenges and strategies for managing digital records in a public organization: Findings from the TEAM China case study. *Proceedings ofInterPARES 3 Symposium*. 4-6 June, pp 243-278. Seoul, South Korea: National Archives of Korea.

Wu, P. H. J. (2009). Towards preservation, description and arrangement of web records: A case study of archiving professional seminar e-learning space.*Proceedings of the InterPARES 3 International Symposium*. Seoul, South Korea: Sungkyunkwan University.

Zawiyah, M. Y. (1999). *Records management in the business community in Malaysia: A study of the understanding of concepts and practice and the development of a model*. University of Wales Aberystwyth, UK.

Zuraidah, A. M. (2010). *A case study on the preservation of digital cultural heritage in Malaysia* (Unpublished Doctoral dissertation). UiTM.