Preservation of Electronic Records Event History Metadata (PEREHM) in Malaysia Government Agencies: Evaluation and Validation

Ap-azli Bunawan, Sharifalillah Nordin, and Haryani Haron
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
40450 Shah Alam, Selangor, Malaysia.
ap-azli@salam.uitm.edu.my, {sharifa, haryani}@tmsk.uitm.edu.my

Abstract — This paper proposes a model for the preservation of electronic record (e-record) event history metadata (PEREHM) as one of the best practices in preserving electronic records in Malaysian Government Agencies. This study was conducted in response to the electronic government (e-government) initiatives which was launched in 1997. The model was designed to ensure the applicability of constructive variables of e-record preservation metadata in government agency transactions. The validation of model and supportive response from National Archives of Malaysia (NAM) as well as the NAM state branch has been made through in-depth discussion on the issues in preserving the electronic records metadata. This is to ensure that the model can be used in delivering the utmost impact to the e-record preservation activities. The paper also discusses the standards and policies applied in government agencies that contribute to the development of the model. In addition, the workflow and structure of the model are also discussed. This is to ensure that stakeholders contributing to the development of this model will not be affected by collaborative aspects, standard of compliance and ethical issues.

Keywords - E-Government; PEREHM Model; E-Records; Preservation; Event history; Metadata

I. INTRODUCTION

The electronic government (e-Government) initiative was launched in Malaysia in the year 1997 by the former Malaysian Prime Minister, Tun Dr Mahathir Mohamad as one of the seven flagships of the Multimedia Super Corridor (MSC) initiative [1]. The e-Government seeks to make interaction easier for the public and businesses by enhancing their accessibility of electronic information and document. An electronic document (e-Document) is defined as information recorded that requires a computer or other electronic device to display, interpret, and process it [2]. The e-document often created according to individual preferences, making it harder to find, use and manage those [3]. On the other hand, the e-Document will be converted to e-Records when business transaction occurred. Meaning that, the e-Document is a daily document whereby the e-Records happen after transaction has been made on the e-Document within agencies. The electronic or digital record may be ‘born digital’ (created using computer technology) or they may be converted into digital form from their original format [4]. Managing the e-Records in the government agencies seems incomplete without having a dedicated system. The dedicated system that currently is being used by the government agencies known as Electronic Records Management System (ERMS). ERMS acts as a platform for managing the e-Records. ERMS provides the technological component of a model for systematic and structured management of records.

Besides that, E-SPARK (Elektronik Sistem Pemeliharaan Arkib Rekod Kerajaan) project was established to ensure that the e-Records generated by the government agencies are preserved while maintaining its authenticity, availability and continuity so that they are need to create these components, incorporating the applicable criteria that follow accessible at any time [5]. The E-SPARK consists of 2 phases; phase 1 and phase 2. The E-SPARK project phase 1 aims to identify policies, guidelines, processes and procedures in preserving e-Records in the government agencies.

In order to generate a continuous support towards e-Records preservation, the E-SPARK project phase 2 has been developed by the National Archive of Malaysia (NAM). In this phase, the focus is more on the dedicated system which are known as Archival Management System (AMS). The objectives of the AMS are to focus on the digital repository for storage, preservation and access of the e-Records that have been transferred by the government agency to the NAM [6]. This is where the NAM needs to ensure that the AMS preserved the e-Records continuously

The NAM has been created in 1st December 1957 as the institution that has the authority to keep and preserve all government records. The NAM is also known as a Government Records Office [7]. The AMS is developed to provide storage space for digital object and its metadata based on Open Archival Information System (OAIS) - Tesella SDB (Safety Deposit Box) [8]. The SDB is designed to preserve vital digital information. SDB provides a security of access for vital digital assets as being required by NAM [9]. The relation between OAIS and Tesella SDB need to be improved in preserving e-Records. This is because the amount of e-Records in government agencies has increased [10].

DOI 10.5013/IJSSST.a.17.32.32 32.1 ISSN: 1473-804x online, 1473-8031 print
II. BACKGROUND OF STUDY

In order for e-government services to gain widespread acceptance, they must have the same standing as well as the confidentiality and security equivalent to the paper process [11]. More importantly, as reiterated by [12], the accountability of records is required in any legal matter that involves the processes of finding, preserving, and producing information in response to a dispute or lawsuit.

On the other hand, the proper understanding on records itself is important. The records act as evidence and thus, the source of information must be authentic and trusted. If a record is to function effectively as evidence of an activity, it must first be compliant with any external requirements in the environment where the organization operates. A records management is an essential element of organization infrastructure together with financial management and human resource management [13]. A systematic record management is important so as to ensure order, efficiency and accountability in business transaction. Thus, without records to provide documentary evidence of its activities, no government or organization can be held accountable [14].

Besides that, the secure storing, description and dissemination of the electronic records as well as archiving of all other records serve as legal certainty [15]. There are many types of electronic records which require proper management to ensure the availability of the records for future use, which will benefit organizations and subsequently the nation. Even though most of the transactions are done electronically, many organizations or agencies are still practicing the hybrid approach, which means everything in the system has to be printed, transferred into a manual file and stored in a file room for future reference.

As a response to this problem, the National Archive of Malaysia (NAM) has developed the National Archive Act 2003 as a guideline for Malaysian Government Agencies to effectively and efficiently manage electronic records. On top of that, structuring the policy in the business transaction activities has become the main focus of the government in producing the best practice approach in preserving e-records. The e-government initiatives have enabled every government agency in Malaysia to be linked with other government or organization can be held accountable [14].

Moreover, according government agency staff, a system that is applied in most departments should comply with the national policy, under the E-SPARK. Unfortunately, there are issues occurred in terms of preserving e-record metadata. The metadata of e-records are not preserved effectively. This includes the loss of e-records event history metadata in daily transaction. As a result, agency faces a lot of problem when submitting the e-record reports to NAM. NAM cannot accept the e-record preservation reports when the metadata are incomplete. Because of these issues, the study would like to present a model that highlights the importance of documenting each event in e-record transactions effectively.

A. Current Issues on e-Records Preservation Metadata – NAM Case Study

Based on the discussion with a staff from the NAM, the AMS is not completely support the process of preserving the e-Records metadata. This is because, the AMS only rely on OAIS model. The OAIS model has limited functionality. The OAIS model only focuses on the workflow of the e-Records during preservation process. It does not concern on the completeness of the metadata. This is where when preserving metadata for the e-Records; it must be supported by various preservation metadata models. Each of the models has its own component which can overcome the lack of metadata content in the AMS [19].

Even though the AMS is the only system that has been accepted by NAM in preserving e-Records, it still has its weaknesses. By focusing only on the version development of the AMS, this has lead to the confusion from the government agencies in capturing e-Records metadata. This is because, The AMS only can receive a metadata that have been filtered by the NAM manually before it can be registered in the system. Filtering the metadata manually means, NAM will use their staff to sort the e-records metadata on the printed version. Then, the staff will do a checking on every metadata for each the e-record. This will...
lead to the possibility of incorrect metadata because of a human error. If the metadata is incorrect, meaning that NAM has preserved wrong metadata in AMS.

As discussions continue, the other critical issues that have been highlighted by the staff are on the event history metadata. These issues occurred when the NAM receive the final version of the e-Records from the government agencies, the metadata on event history is incomplete. Meaning that, the e-Records metadata that has been finalized by the government agencies is not comply with the existing metadata model. In this case, the NAM faces difficulties in registering the e-Records metadata on the AMS for the preservation purposes. As a result, the NAM needs to return the e-Records and its metadata to the government agencies for the correction purposes.

B. Current Issues on e-Records Preservation Metadata – Government Agencies

As a respond to this issue, a discussion session with the staffs from government agencies have been conducted. As government agencies that involve in managing e-Records, they need a dedicated model that can be followed on documenting each of the e-records metadata [20]. By having this standard on capturing the e-Records metadata, they faced a problem in identifying metadata needs to be preserved. This is because, current model that being used is too structured. Thus, most of the standard structure has been design for those who have foundation in managing e-Records. Otherwise, for those that do not have the knowledge in managing e-Records, they will have difficulties in considering the standard. These difficulties have lead to the issues on incomplete metadata reported.

Then, the staff also addresses issues in finalizing the event metadata for e-Records. This means, they explain on the business process of creating a metadata for e-records. The issues begin when e-Records are being modified from one department to another; the event of the e-Records has not completely been documented. This is where they have lost the e-Records event metadata. In some cases, they only report the final version of the e-Records metadata and not for the overall event on the e-Records metadata. As a result, the government agencies need to resubmit their report to NAM repeatedly.

On the other hand, the staff also recommended the development of one dedicated model specifically for the preservation of the e-Records metadata. This model will standardize the entire process and event of the e-Records metadata in government agencies. By having this model, it is also easier for the government agencies to understand the actual process in documenting event on the e-Records metadata. This will help the staff to produce complete report on the e-Records metadata and in line with NAM requirement.

In most country, archives need to document any action taken to the object so that in the future, users can judge whether an object has been altered and, if so, in what way, by what process, when and by whom [21]. This process is also known as an Event History Metadata [20]. The event history metadata has been defined as a documenting the trail of past records, events or other action on both entity and its metadata [20]. Thus, event history metadata has been created to ensure the metadata of e-Records is recognized and valid for future reference. As for government agencies, early capture on e-records creation metadata should be made in order to safeguard their authenticity, reliability and usability [22].

In this situation, government agencies need to move on and severely go for the e-Records preservation activities. If not, e-Records metadata would be lost [23]. In avoiding these negative impacts happen, government agencies have decided to follow the current practices from other country [24]. As a result, the preservation process becomes worst and difficult. This is because, the content of e-Records metadata is different in each countries. The differences have lead to the issues in preserving metadata for e-records when there is no complete preservation metadata model to be referred. This is where issues on event history have occurred while preserving e-Records metadata.

III. PROBLEM STATEMENT

As being discussed above, the e-records event history metadata preservation is the main issues that need to be encountered in this study. This is because the unavailability of complete e-records preservation metadata model that comprises multiple components does not exist in the government agencies. This is where it is highly recommended for the development of a complete preservation metadata model that includes the proper management of e-records metadata, as well as event history metadata. These models can be fully utilized by the government agencies and will directly adequate them to preserve the e-Records metadata and the e-records event history metadata properly. Based on this drawback, this study is designed to address the issues regarding event history metadata as well as formulating complete e-records event history metadata preservation model.

IV. PRESERVATION METADATA MODEL

As being mentioned before, the e-Record is part of the business transaction process in the government agencies. Each and every process in preserving the e-Records, the metadata component should be preserved as well. This is because, the metadata become a standard component of most electronic preservation implementations [25]. Metadata can be defined as data about data which also known as information about data [26]. The metadata is structured information. It is divided into three (3) categories [27]. There are;

- Descriptive metadata - It describes a resource for purposes such as discovery and identification. It
can include elements such as title, abstract, author, and keywords.

- Structural metadata - It indicates how compound objects are put together, for example, how pages are ordered to form chapters.
- Administrative metadata - It provides information to help manage a resource, such as when and how it was created, file type and other technical information, and who can access it. A preservation metadata is a subset of administrative metadata. The preservation metadata contains information needed to archive and preserve a resource.

A. Open Archival Information System (OAIS)

The model comprises the information package as a conceptualization of the structure of information and consists of an object to be preserved and the metadata necessary for its long-term preservation and access which was bound into a single logical information package. Three types of information objects exist include the submission information package (SIP), the archival information package (AIP) and the dissemination information package (DIP) [29].

The Archival Information Package (AIP) is the focus of preservation activities. It must contain the complete set of metadata necessary to support long-term preservation and access. Four types of information make up the AIP: Content Information, Preservation Description Information, Packaging Information, and Descriptive Information. The first two contain metadata directly relevant to preservation. Hence, the Descriptive Information, while certainly required by the repository, is considered more relevant to resource discovery than to the preservation function itself.

Besides that, metadata is necessary for the user to understand the objects that are bound into the Content of Data Object. This also leads to the finalizing the Representation Information in the case of digital objects, to the string of bits that comprise the object which is also referred to as the Content Information [29]. The metadata necessary to carry out preservation functions are found in the Preservation Description Information. It can be categorized into three types of information in general:

- Reference information – uniquely identifies the Content Information,
- Provenance information – details the history, documents any alterations to content and form over time, and chain of custody information,
- Context information – documents the relationship(s) of the Content Information to other Content Information objects.

Currently, government agencies used OAIS models in preserving the e-Records metadata [5]. The OAIS model has been defined as people or organization that accepted the responsibility to preserve information and make it available for the community [29]. Despite, OAIS is receiving information or the e-Records from the people or organization along with associated metadata [30]. This type of metadata is depending on the report that has been provided by the people or organization in the registration process. On the other hand, OAIS is not only specific for the e-Records because OAIS is practicing a flexible implementation [31]. Generally, OAIS is difficult to be used because of its complexity. The OAIS environment comprises of three distinct external entities which are Management, Producer, and Consumer [32].

In this study, the focus is on the “Producer” in External Variable on OAIS. This is because a “Producer” is referring to the government agencies [5]. Government agencies need to submit a Submission Information Package (SIP) to the NAM in ensuring the e-Records metadata can be preserved properly. In the SIP, it does contain overall metadata on e-Records. The metadata on e-Records should be completed and packaged according to NAM guidelines and procedures. But, the issues will arise if the metadata that being packaged is not complete as per request. This has lead to incomplete metadata submission to the NAM.
B. Preservation Metadata: Implementation Strategies (PREMIS)

The Preservation Metadata: Implementation Strategies (PREMIS) is one of the requirement and guidelines [33] that need to be considered in preserving the e-Records. PREMIS highlighted the entire requirement in preserving digital records metadata in order to ensure there is no issues occurred in the future. PREMIS has defining the elements of preservation metadata and supporting implementation in digital repositories [34]. The PREMIS data model has five primary types of entity which are intellectual entities, objects, events, agents, and rights [35].


The government agencies must follow a metadata model that includes six (6) broad grouping of metadata that need to be included in managing e-Records metadata. There are Identity, Description, Use, Event Plan, Event History and Relation. Each of this group has their own attribute and component which has been used as guidelines in managing e-Records metadata in the government agencies. Thus, this model is important because it covers the entire aspect of e-Records metadata. This includes from the creation of e-Record until it being transferred to NAM.

D. Encoded Archival Description (EAD)

EAD is a metadata schema for archiving digital resources [28]. In addition to the content description of digital resources, EAD has elements for structural description [28]. The EAD is the heart of the electronic finding aid, providing the descriptive tags for the intellectual structure of the collection. It is a “set of rules for designating the intellectual and physical parts of archival finding aids so that the information contained therein can be searched, retrieved, displayed, and exchanged in a predictable platform independent manner” [36]. EAD consist many variables in order to evaluate the e-Records preservation metadata [37]. But in the process of evaluating the e-Records event history metadata preservation, there are seven (7) variables that have been chosen by the expert [6].

V. PROPOSED THEORETICAL MODEL

In providing the solution to the issues, this proposed theoretical model was being developed. The development of this model is to enhance the OAIS model in terms of its Producer attributes. As being discussed above, the Producer attributes has lead to the incomplete metadata given to e-records preservation purposes. Therefore, by underlying the list of variables on e-record event history metadata, it can give a solution to the government agencies before they send their final version of e-record metadata to the Archive.

The added value or enhancement on OAIS model including structuring the types of metadata that being used by government agencies in order to achieve the objectives of preserving the e-record metadata for evidence and security purposes. By reviewing the PREMIS, ISO/TS 23081-2 and EAD models, it has been validated by the expert from NAM in terms of the completeness or best practices in managing the e-records metadata in Malaysia government agencies. Each of these models has contributed their aspects of preserving the event history metadata. The details of contribution are shown in the Table I.
TABLE II.
IDENTIFIED VARIABLES FOR A MODEL IN PRESERVING E-RECORDS EVENT HISTORY METADATA

<table>
<thead>
<tr>
<th>No.</th>
<th>List of Variables Contribution from PREMIS, ISO/TS 23081 and EAD</th>
<th>Contribution from</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Event Title (EAD)</td>
<td>PREMIS, ISO/TS 23081 and EAD</td>
</tr>
<tr>
<td>2</td>
<td>Event Identifier (PREMIS)</td>
<td>PREMIS, ISO/TS 23081 and EAD</td>
</tr>
<tr>
<td>3</td>
<td>Event Type (PREMIS)</td>
<td>PREMIS, ISO/TS 23081 and EAD</td>
</tr>
<tr>
<td>4</td>
<td>Event Date/Time (PREMIS)</td>
<td>PREMIS, ISO/TS 23081 and EAD</td>
</tr>
<tr>
<td>5</td>
<td>Event Detail (PREMIS)</td>
<td>PREMIS, ISO/TS 23081 and EAD</td>
</tr>
<tr>
<td>6</td>
<td>Event Outcome Information (PREMIS)</td>
<td>PREMIS, ISO/TS 23081 and EAD</td>
</tr>
<tr>
<td>7</td>
<td>Event Relation (ISO/TS 23081)</td>
<td>PREMIS, ISO/TS 23081 and EAD</td>
</tr>
<tr>
<td>8</td>
<td>Event Trigger (ISO/TS 23081)</td>
<td>PREMIS, ISO/TS 23081 and EAD</td>
</tr>
<tr>
<td>9</td>
<td>Maintenance History (EAD)</td>
<td>PREMIS, ISO/TS 23081 and EAD</td>
</tr>
<tr>
<td>10</td>
<td>Maintenance Event (EAD)</td>
<td>PREMIS, ISO/TS 23081 and EAD</td>
</tr>
</tbody>
</table>

Based on the reviews of the models, there are Ten (10) variables that being identified in this study. These variables were based on the combination of event history variables from PREMIS, ISO/TS 23081-2 and EAD model. In ensuring this model has its originality, the variables have been validated by the experts from National Archives of Malaysia. This proposed model also can be addressed as a PEREHM (preservation e-records event history metadata).

VI. PEREHM MODEL FUNCTIONAL VARIABLES

In brief, PEREHM model contributes significantly to each activity in the government agencies. Testing and evaluation are carried out by a group of experts in related field. In the aspect of testing, it is unethical to reveal the details of the people who are involved in the validation process. The PEREHM model consists of the character and metadata that are connected with the current practices on e-record. This has been proven by generating the key element in each of the variables included in the model. Thus, the workflow of the records creation is based on the current practice in Malaysian Government Agencies as being mentioned in the previous section.

A. Event Title

An event title is based on the title that is used as a metadata in the records creation. This means that any changes of event that are amended in the system must be recorded. This is where the “Event Title” variables are important in order to notify the final authority of the e-records.

B. Event Identifier

An event identifier is the key aspect that must be considered when there are changes made towards the e-records metadata. It includes the “Identification” on any event that happens in the process of managing the e-records.

C. Event Type

An event type is also important in order to identify the type of event that happens towards the changes on e-record. The event type is needed for documenting the “type of modification” that has been made in e-record.

D. Event Date/Time

An event date/time is the most significant variable that is important in ensuring the transaction of e-record is captured effectively. This includes the “Day/Month/Year” as well as the “Hour/Minutes/Second” of the event activities in the e-record system.

E. Event Details

An event detail is highly connected with selected variables in PEREHM model. The details of the event are important in order to ensure proper manner of e-record reporting. It is because if an event is justified in brief, it can lead to the mismanagement of the e-record final report.

F. Event Outcome Information

Event outcome information contributes to the overall aspect of the event. This is also means that the impact of the event can be identified when there is an “Outcome Information” being embedded in the e-record system. The outcome information on the event reveals that the purpose of e-record is needed for modification.

G. Event Relation

Most of the events that happen in e-record are related to each other. For an example, when there is a change in the subject matter, it will also affect the record number of the e-record. This type of event must be documented in order to identify the relationship between two or more events that have been made in e-record.

H. Event Trigger

An event trigger is one of the significant variables in PEREHM model. This is because the trigger factor that is embedded in the e-record metadata must have its own trigger point. An example is the closing date. The life of e-record is based on its function and value. Once the e-record meets its deadline or date for disposition, a few metadata will automatically change. The event of e-record happens based on the trigger point that is selected during the creation of the e-record.

I. Maintenance History

Each event or history of the e-record must be maintained until it is submitted through SIP to NAM. This history acts as an evidence of the transaction and must be included in the final submission for the purpose of permanent preservation. Thus, this history is able to describe the origin of the e-record until up to the latest condition.
J. Maintenance Event

In ensuring each of the event details is preserved effectively, the maintenance event variable is needed. This is because when there is an event added to the e-record, it can be deleted, change or removed at liberty. This is where the maintenance of every event in e-record is compulsory in order to prevent any mishandling of e-record in the organization.

VII. PROPOSED MODEL VALIDATION

The validation of PEREHM model has been made at several NAM state branches and headquarters in Malaysia in order to give a maximum impact on the model. This is because it is the lead agency in each state which is responsible in preserving the e-record in Malaysia. As this study is conducted throughout Malaysian government agencies, it is a must to have the approval from this lead agency which acts as an expert.

Hence, the findings show the supportive effort and recommendation in ensuring that the model is in line with the current practices in government agencies. According to these experts, it is important to have this PEREHM model in order to make the e-record preservation synchronized with the standard and practices that have been conducted by NAM.

Thus, they decisively declared that in preserving the event history on e-records metadata in Malaysian government agencies, it is important to avoid the occurrence of any miscellaneous when submitting the final report for permanent preservation in AMS. They acknowledge the use of PEREHM model to document any transaction of the e-record for evidence and security purposes. The details of their response are shown in Table III.

The validation of the proposed PEREHM model has been completed by conducting an in depth discussion with staff from NAM headquarters and branches. This process includes the validation on the selected theoretical model which includes PREMIS, ISO/TS 230781 and EAD. In these early stages of validation, they conform that 10 selected variables in PEREHM model are an important to overcome the issues in preserving the e-records event history metadata. They are approved that, 10 selected variables on event history metadata should be captured since in the early process of creating the e-records in government agencies. As for now, there five (5) NAM has validated this PEREHM model.

VIII. FINDING AND DISCUSSION

Based on the finding of this study, it is concluded that the main aim of this study is not to compare the features and function of the existing system. It is more to add a significant value on the best practice of preserving the e-records event history metadata in Malaysian government agencies. Thus, by having an in-depth discussion with the experts, it generates more ideas on the actual issues that occur in the Malaysia government agencies.

Hence, the processes of validation have exemplified the needs of having the proper model on e-records event history metadata. This is because most government agencies still depend on the system features and function and not on the principle of preserving the e-records as a whole process. Therefore, the study presents the validation of an existing theoretical model which includes OAIS, PREMIS, ISO/TS 230781 and EAD.

This type of model is widely used in most archives in the world. Therefore, in the case of Malaysian government agencies, it should be embedded in the system process flow in order to obtain the maximum output of the e-record preservation activities. Moreover, this type of model implies the best practice in most countries that have completed the process of preserving their e-record metadata effectively. On top of that, by adding certain value to an External Variables in OAIS model (Producer), it will positively augment the OAIS model to be applicable in the case of preserving the e-records metadata in Malaysian government agencies. Since an event history indicates ten variables in PEREHM model, it shows a significant approach has been made through tremendous discussion and validation by experts.

IX. CONCLUSION

As a conclusion, managing e-record in Malaysian government agencies is important in order to generate the contextual and structural benefits to the preservation activities. There are many systems that are used by the government agencies but those ensuring its usefulness and compliance with the standard on e-record are limited. There are a few systems that comply with e-record standard and procedures. However, those that meet a highly recommended system that contains all function on e-records are also limited. By developing the PEREHM model, all these can be embedded in e-record system in Malaysian government agencies. This is because the PEREHM model includes ten variables that are most significant to be implemented and applied in Malaysia government agencies.
ACKNOWLEDGEMENT

This study is part of an ongoing research for Doctor of Philosophy Computer Sciences in Universiti Teknologi MARA. This research was sponsored by Ministry of Education under SLAB/SLAI government servant scholarship. This sponsorship is effective from September 2014 until September 2017.

REFERENCES


