Database design of the Malaysia public figures web archive repository: a social and cultural heritage web collections

Farrah Diana Saiful Bahry and Noraizan Amran
Faculty of Information Management, Universiti Teknologi MARA Selangor Branch, Puncak Perdana Campus, Malaysia
Tesa Eranti Putri
Faculty of Vocational Studies, Universitas Airlangga, Surabaya, Indonesia, and
Muhammad Idzwan Ramli
Ministry of Tourism, Arts and Culture Malaysia, National Archives of Malaysia, Kuala Lumpur, Malaysia

Abstract

Purpose – The growth of web emerging technology and data visual demand from the World Wide Web (WWW) makes the need for information repositories become vital. Proper database development will assure the repository managing web content effectively aligns with web archive metadata standards. This paper aims to present the database design process for web archive content repository specifically to maintain social and cultural heritage values upon Malaysians as Mfigures.

Design/methodology/approach – The empirical process start with literature review and validation from expert on the elements and scopes of research. Then, structured database design guideline which part of database life cycle (DBLC) was applied and combined with the step of comparative and mapping the conceptual model with metadata standard that is relevant with web archive content. The paper focuses on the first three stages: Database Initial Study, web archiving and Metadata standard mapping; and conceptual design to focus on data modelling. Another two stages of database design are logical design and physical design will be exposed later.

Findings – The empirical process has produced initial conceptual data model, database structure that can be a basis of web archiving repository. The data model had also been verified with metadata data standards to assure the database structure implementation cater the need of web archiving repository features especially web information discovery.

Research limitations/implications – Nevertheless, database design is the most effective way to develop good information architecture on the Net, but the absence of some important fields on related tables have been identified such as subject, language, coverage, right, publisher and contributor. The MFigures' database schema will continuously improve for better scope and coverage of web archive content suite with future information demands on the WWW.

Practical implications – The conceptual data model act as a communication tool by the technical team in web application development. It can be revisited to suite with other different database management system or to suite with other similar scope of information repository requirements.

Social implications – Mfigures was uniquely designed for collecting Malaysian social and cultural heritage, which are rarely design before, and it can be beneficial as Malaysia society future references for excellent motivations roles and successful stories.

Originality/value – The Mfigure conceptual data model was empirically design and gone through a proper validation process by the industrial and academic experts.

Keywords Web archive repository, Metadata, Web preservation, Database design, Social and cultural heritage, Malaysia

Paper type Research paper

Introduction

Web archiving is used to create a copy of an updated version of the website for that particular time for the purpose of long-term web preservation and retrieval in the future. Web content can be considered as virtual cultural heritage or cultural heritage information resources. Capturing positive people's experiences and their ways of living toward success could bring multiple thoughts to others. It also becomes part of social and cultural memory to be kept and recognized in the future. Most stories related to people, environment phenomena or any live scenarios on earth happened have been captured through webs or online. Nevertheless, the lifetime of each web document also

The current issue and full text archive of this journal is available on Emerald Insight at: **https://www.emerald.com/insight/2514-9326.htm**

Collection and Curation
41/4 (2022) 133–140
 [ISSN 2514-9326]
[DOI 10.1108/CC-09-2021-0025]

The authors wish to thanks to colleague and superiors in the faculty Information Management, Universiti Teknologi MARA Selangor, National Archives of Malaysia and Faculty of Vocational Studies, Universitas Airlangga for giving us opportunity doing the joint collaboration in teaching and research project until the completions.

Received 28 September 2021
Revised 20 December 2021
Accepted 26 December 2021

cannot be predicted and it can be easily inaccessible at one age of time.

According to UNESCO, cultural heritage encompasses several main categories which are tangible, intangible, natural and event-based. Part of tangible cultural heritage are websites that contain cultural heritage values. Intangible cultural heritage can be transcribed as acceptable and recognizable ways of living and practices by the community that are inherited from the past and continue to evolve day to day. It can be a custom, knowledge of tradition and ways of doing or skills transferable from one generation to another or blended in multicultural society. Whereas, digital heritage is defined as computer-based materials of enduring value that should be kept for future generations (UNESCO, 2021). In any category, Cultural Heritage Information (CHI) nowadays is generated in social media platforms or websites. The meaningful web content should be revealed and maintained either from digital objects such as texts or even from databases, still and moving images, audio, graphics, software and web pages or another increasing new web file format.

Web archive repository is a central file storage location, which has versioning control and facets on sustainable, trusted and organization of web archive content to function properly (Taylor, 2017; Momin and Gaonkar, 2016). A relational database, such as Oracle, Mongo DB, Cassandra, PostgreSQL or MySQL, is used to run most repositories. Data is arranged into tables in a relationship database, which are linked together by relationships. Database such as Oracle, DB2 and Microsoft SQL Server are proprietary systems that need to be licensed while MySQL, MongoDB and Postgres are open-source database management systems that are mainly born as web data storage for web-based data structure and application platform. It also displays options for depositing unstructured web archive content and description such as URLs, timestamps, descriptions, sources, content type, format, date changed and date validated. To function properly, the repositories must be long-term, dependable, well-supported and controlled. To prepare and make available the semantic meaning of intangible cultural heritage information gathered from online environments, Sugimoto *et al.* (2018) stressed out the importance of data models in defining metadata schemas. Metadata mapping is also crucial for assuring data aggregation can be accomplished smoothly. Developing web archiving repositories for special collections or single subject matter need to include contextual information for metadata mapping to ensure meaningful content retrieval in the future.

In designing web archiving repository data models or information models, it is good if proper data design can be instigated or framed. Web archive repositories nowadays emerge to suit specific fields of knowledge collection and discovery. Thus, the conventional but yet applicable data model guidelines were applied especially for students' understanding who become a novice information researcher and web curator. The data model design guidelines consist of a few stages that need to be followed such as formulate data requirements, its relationships and define it as business rules; identify basic building block of the data model; visualize the conceptual model such as in entity relationship diagram and conceptual ERD; and then map it with acceptable web archive metadata standard. The guideline has become a methodology in designing the relational database concept of web archive repository. In the next section, we state and demonstrate the guideline as well as instructions on how to obtain the information. The aim of this task is to have a better understanding on how a data model is conceptualized and map with descriptive and preservation metadata conforming to a granularity set of data elements (Dooley and Bowers, 2018). The next sections of the article follow steps in designing a data model for database creation.

Literature review

Understanding of web archive and web archive repository

Web archiving process normally is a recurring process, which captures any web content update by the web document such as a website. Web archiving strategies can be categorized by site-centric archiving or topical domain-centric archiving (Khan and Rahman, 2019; Brügger and Laursen, 2019). Site-centric archiving is mostly done by corporate bodies, institutions or even individuals for intellectual collection building purposes. Topical domain-centric archiving is driven by research needs with the purpose of collecting web documents for further analysis or content retrieval. The essential component in the web archiving process is the authorization. Authentication and versioning of the web document need to be well acknowledged and protected. Due to the short lifespan of digital objects on the web, the need for web archiving is crucial. As an outcome, multiple versions of the same web content need to be well organized and maintain their semantic values. Thus, an effective web archive repository is vital especially for capturing valuable content that is related to Malaysia cultural and heritage web content (Schafer and Winters, 2021). Malaysian initiatives towards web archiving still remain scarce, but Malaysia has reached 29.03 million people accessing the internet and becoming the content creator especially in social media.

There are several problems identified in collecting and preserving the web content as it has too much similar, unvalidated and irrelevant information found on the internet. It is hard to identify which are the most relevant and authorized websites because the creator has not stated their personal information on their websites. Therefore, this problem can cause the reduction of information's qualities in the websites. Filtering similar website content also will be time consuming because there are thousands of similar websites contents will be found when performing topical or subject matter of web searching. In terms of preserving cultural and heritage value of web content, keeping the authorized and legitimate website in the web archive repository can securely protect its existing nature of web format, content's ownership and semantic values.

In developing a web archive repository, detailed descriptions in the web archive process are crucial to be captured such as archive date and time, website owner, web content, website last update. To effectively access and retrieve in the future, semantic value of each web content either in visual format, textual or other media need to be well tagged or captured as well. Here, the role of metadata is important to keep the heterogeneity throughout most platforms on the web

(Vlassenroot *et al.*, 2021). Advantage of the new advanced information technology such as semantic web, crowdsourcing, information visualization, text mining, social network analysis and ontology development, means each web archive content element also can be analysed according to each segment and perspectives (Németh and Drotos, 2019).

Metadata in web archive, social and cultural heritage

Metadata is commonly known as data about data, which is structured information that describes or explains the data or information resources contained in any database or storage (Eichler *et al.*, 2020). Metadata standard such as Dublin Core (DC) leverages the descriptions of information resources into several categories, The first category is content description such as title, subject, description, source, language, relation, coverage. While the second category maintains the web archive content's intellectual properties and rights, such as creator, publisher, contributor and rights disclaimer. The third category, DC metadata standard, also encapsulates the structural manifestation of web archive content such as date, type, format and identifier.

In an information system's information organization or database tools, metadata expresses data properties and its conceptual relationships between uniquely recognized conditions (Talib and Osman, 2020). It may be used in much more sophisticated ways, allowing logic and inference to play a part in the search process. Such features can boost the precision and recall of browsing, which are crucial for discovering appropriate details amid the massive volumes of data on digital resource objects. Table 1 shows some metadata creation for cultural and heritage collection development.

The growth of knowledge creation on the World Wide Web has made the role of metadata formation more crucial. Thus, the process of metadata modelling has become a precedent in which it refers to a type of meta modelling used for the analysis and construction of information models to become structured (Tomasi, 2018). In Web archiving, web archivists practice to prepare metadata on the web archive content and files which are captured from the web, and the content itself for assuring web preservation can be implemented (Ayala, 2020). Di Pretoro and Geeraert (2019) explained that metadata practices varied from one web archive initiative to another and their PROMISE project achieved the inclusion of metadata into a single web archive file format which in warc file.

Objectives

1 to design web archiving database repository which suits the descriptive metadata standard aiming for social cultural heritage; and
2 to identify lack of metadata element for knowledge discovery on Web archiving repository of Malaysia social and cultural heritage (Mfigures).

Scope

In this project, researchers aim to design a web archive repository database as a storage platform embedded with descriptive and preservation metadata related to Malaysian social and cultural heritage. The goal in this project is to identify the extent/level of detail (LOD) appropriate for preserving Malaysia excellence figures in terms of their achievement, experience and strategies to success. The project's outcome was named as MFigures and the data structure includes a metadata element that enables possible expansion and dissemination of the web content which aims for a virtual social and cultural heritage web repository.

MFigures web Archive repository has a role to manage and control version of intangible cultural heritage web content related to public figures in Malaysia. MFigures plan to deposit all web archive content of Malaysian well-known public figures, including honorable sportsmen, millionaires, outstanding researchers in various professions, respectable politicians and many others. Public figures' stories, speeches, events involvement, daily lives' acts and thoughts can have a positive vibe to influence the young generation in the future. All of this stuff is significant in terms of representing Malaysia's social and cultural significance. In the current IR 4.0 and upcoming era, web content has become major sources of information in peoples' life, thus semantic descriptors of each web content must be tagged for valid judgement and ease of understanding especially when it comes to social beliefs and role model influences. Having a central repository of public figures' stories and their related web content also can become valuable social and cultural heritage when their achievement, recognition and contribution story can be retrievable after decades of time. Thus, the web archiving process which starts today is producing web archive content that needs to be deposited and maintained in MFigures web archive repository. The web archiving process used Httrack and Conifer tools. Both tools generate two types of files: a warc file and a real-nature web document. To assure information completeness of the web archive content, both types of file need to be deposited in the MFigures. Thus, each web archive content probably has two types of files to be kept and linked together.

Database creation methodology

According to Coronel and Morris (2017), Database Life Cycle (DBLC) consists of six stages including database initial study, database design, analysis, implementation, maintenance & evaluation and operation and evolution. In the database design stage, the steps involve conceptual design, logical design and physical design. The specific method used to design the web archive repository is shown in Figure 1.

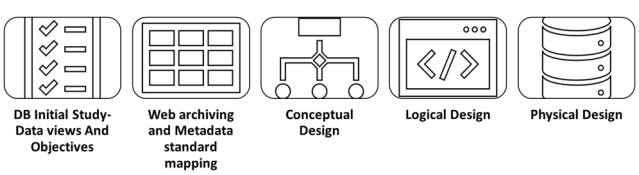
The first step is to identify, analyse and refine business rules. It entailed researching company activities, gathering data and identifying data items by identifying the following: needs, users, sources and constitution. According to Coronel and Morris (2017), a business rule is a clear, concise and unambiguous definition of a policy, practice, or principle inside a company. Database designers need to understand the use of the database in organization, communities or targeting specific users in an online environment. A complete and concise outcome of this step is listed out in specified business rule format in bi-directional statements that relate the two objects or entities within targeted relationships or specified connection behaviors.

Then, the second step here includes to suit the purpose of web archiving repository functionality that is web

Table 1 Metadata creation related to social and cultural heritage collection

Authors/Year	Web Archive project	Metadata included	Metadata adoption
Tan *et al.* (2019)	e-WARISAN virtual heritage (VH)	What: Heritage ID, Title, Heritage Type, Heritage Period and Heritage Time Span • Why: Purpose (reason recorded) • How: Recording Device Parameters, Secondary Device, Environmental Conditions • Whom: Submitter and Date of Submission, Rights Given, Author (Copyright Holder), Sponsor (client) • When: Date (of recording and manipulation) • Where: Location (Latitude/Longitude and direction)	UNESCO's World Heritage metadata
Dooley and Bowers (2018)	OCLC research partnership of Web Archiving Metadata (WAM) Working group's 14 recommended data elements	Collector, Extent, Source of description, Contributor, Genre/ Form, Subject, Creator, Language, Title, Date, Relation, URL, Description, Rights	After reviewed descriptive metadata standard (neutral standard)
Wijesundara and Sugimoto (2018)	Cultural Heritage in Digital Environment (CHDE)	Cultural Heritage Objects (CHOs): individual object information (i.e. item-centric information), whereas non-memory institutions provide CHI in a large granule, such as descriptions about historical and cultural contexts (e.g. the historical background of a CHO)	Dublin Core Metadata Initiatives (DCMI)
Giannoulakis *et al.* (2018)	Intangible Cultural heritage of Folks Dance: Digital Choreographic Model	Dance (Title, History, Country/Region of Origin, Time of origin, Dance variations, Music motif, Lyrics and Dance annotation) Recording (Title, Description, Location in GPS format and human readable, Number of sensors, sensor type and characteristics Calibration and Software) 3D Environment (Venue description, Objects and their description, Description of the lightning) Dancer (Gender, Face Expression, Traditional costume description)	FRBR schema
Amin *et al.* (2012)	IR-BUDAYA	Malay Intangible Cultural Heritage: Oral Tradition and Expression, Knowledge Traditional Craftsmanship, Performing Arts, Knowledge and practices concerning to Nature and the universe; Social Practice Rituals and Festive Events and Performing Arts	UNESCO guidelines
Irfan *et al.* (2018)	ICH Inventory of Intangible Heritage in Penang Malaysia	cultural heritage: material (buildings, objects, museum collections), intangible (know-how, cultural practices) and natural (cultural landscapes) what (the presentation of the theme either a cultural practice, a site, a building, a landscape, a collection, an object or people); how is being valued by communities in present time, and its uses; what is the historical background; which changes have occurred how and why; which actions were implemented concerning patrimonialization	UNESCO guidelines
Musa *et al.* (2020)	Digital archive management for heritage buildings and monuments in Melaka using the Geographic Information System (GIS) technology	Melaka heritage building's structure description: Heritage_Building, River, Road, Cadastral_Lot, JUPEM_Lot, Buffer_Zone, Land_use, Basic_Map, Inventory _Form, Measured_drawing	UNESCO guidelines

Figure 1 Database creation guide for MFigures web archive repository



archiving metadata standard mapping. Existing web archiving metadata standard, social and cultural heritage related metadata and other derivation of metadata model creations have been benchmarked and mapped to include as data elements in MFigures WA repository's structure. The process here was targeting to loop and keep updates whichever new related metadata scheme, metadata element

and standard evolve and meet the expandability of the MFigures WA repository functionalities.

The third step, conceptual design, is actually the first design process that focuses on creating database structure. The conceptual design components mainly are data entities, attributes, relationships and constraints of each business domain specifically in this context are managing web archive contents and administering its version control and future discovery. The output of conceptual design is conceptual data model which is commonly visualized in either Entity Relationship diagram (Ayadi *et al.*, 2021; Amran *et al.*, 2018) or UML class diagram (Tomasi, 2018; Lendák *et al.*, 2020). Using ER diagrams, lowest data abstraction level and granularity can be achieved to capture semantic meaning of each content in the web.

The data model can be translated into database logical design which defines each data format and components in logical structures. Logical model consists of tables, fields of table and relationships mediated among tables. The paper limits the scope of database design and abstraction of entity level until revealing the data values without including the physical design of the database repository.

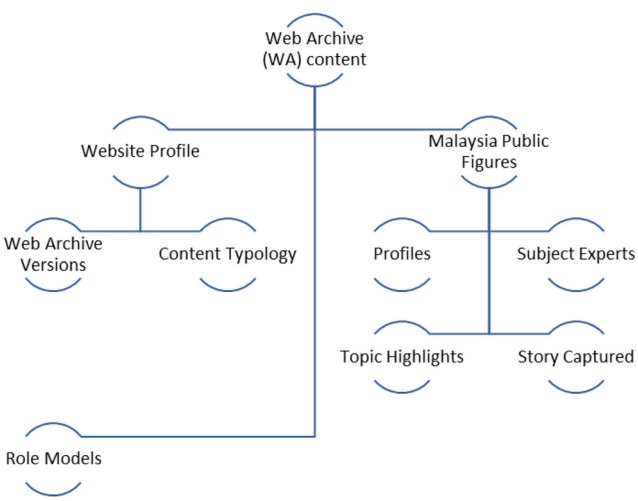
Findings of the MFigures web archive database design

Database design for MFigures web archive repository was empirically conceptualized. MFigures database was designed to store the description of the websites that need to be archived, each detailed process of the web archive that took place for all related websites, versions of the same websites being archived, Malaysian public figures details and their field of achievements. The web content scope that plans to discover and store from the World Wide Web are shown in Figure 2.

Related business rules consolidated under the creation of the MFigures database are:

- a website can be archived many times and each archive belong to each particular website;
- each public figure can exist in many web archive versions and each web archive version is considered as written about one public figure; and
- each public figure can become many good roles and each role can be represented by many figures or names.

Figure 2 Content included in MFigures Web Archive



The appropriate business rules statement which upon user's requirement and database designers' mutual endorsement will assure good association between entities, relevant attributes within the table and valid data restriction or constraints be formed during implementation. The Website archive, Archive content, Public figure, Role and Role model were identified as the entities, which then were transformed as tables. Each entity or table includes its own attributes which are translated as table's fields during database implementation. Entity such as web archive includes detail about web archive process such as targeted website name or title, and website owner. Besides that, the archive content entity keeps versions of the web content creation, which includes date and time creation, duration, file size, no of files and the archive file itself. Other entities also include only necessary details as possible as the initial database design of MFigures.

The relationship between entity website archive and archive content were determined as each website archive has many archive contents (1:M) such as shown in Figure 3.

The relationship between entity website archive and figure was determined as each figure can exist in many websites' archives but each website archive only represents one Figure (1:M) such as a Figure 4.

Last relationship created between figure, role and role figure was many to many relationship from the business rule. Each Role can be represented by many Figures and each Figure can

Figure 3 Relationship between website archive and archive content

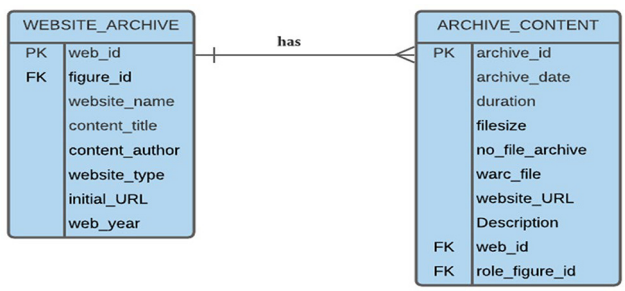
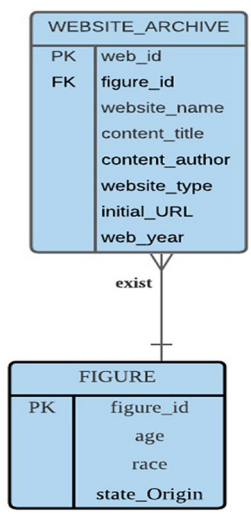


Figure 4 Relationship between website archive and role figure (figure)



contribute to become many role such as an artist not only shows her role as singer but also sometimes she is a good dancer or may become a good businessman at the same time.

Figure 5 shows the external view of ERD for Figure and Role relationship. The common rule or guide followed for many to many (M:N) relationship implementation is the creation of an associative entity or bridge entity or composite entity that took place to relate both entities and simply named as Role Figure entity. Another common rule or guide is that database design is also an iterative process. Once revision has been made on the database structure, either adding new fields in table or combining two tables or even updating relationship type, the database design should also be updated and reported. Database design outcomes such as in ERD or Use Case Diagram are the good communication tools between developer teams within the implementation stage or used by the database administrators during database maintenance & evaluation stage.

The complete initial ER diagram or likely called conceptual model is the full view of representation of the coming database structure. The conceptual model using ER diagrams that describes in depth data values resides in prolific website content such as in Figure 6, which shows number of entities, attributes in each entity and relationships that link between entities. MFigures repository not only aims to deposit the web archive file and URL but also become a knowledge discovery tool in the future related with social and culture depicted from Malaysia's role figures such as distinguished researchers, astronomers, singers, millionaires, honorable activists and politicians.

Then, the benchmark on MFigures fields with metadata standard had to go through some revision to make sure sufficient embedded metadata was included in describing and preserving the web archive content. Table 2 shows that, out of 15, 11 metadata elements related to web archiving applied on the MFigures database structure. Dublin Core metadata and OCLC elements were among the widely used metadata

Figure 5 Relationship between figure and role

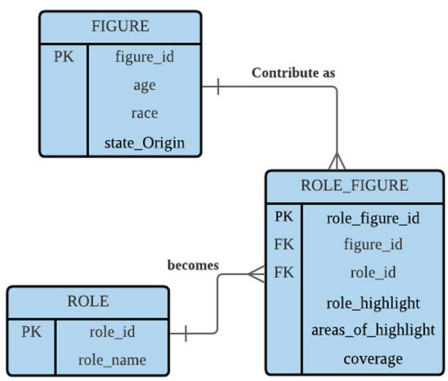


Figure 6 Conceptual model of MFigures WA repository using ER diagram

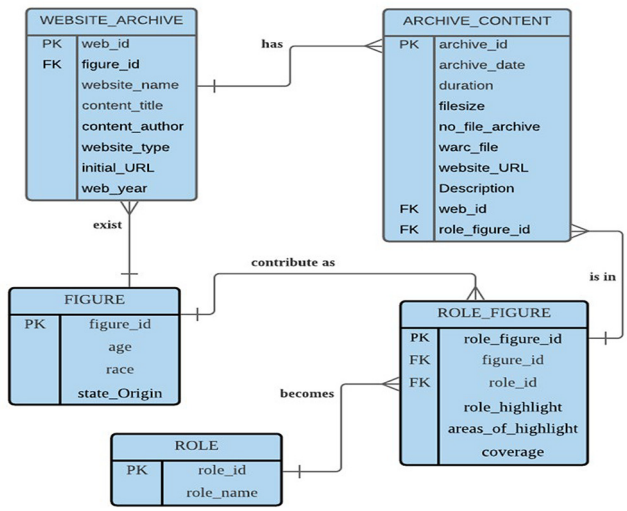


Table 2 Mapping of MFigures fields with web archiving metadata standard

No.	MFigures metadata	Dublin core metadata element	OCLC metadata
1.	Content title	Title	Title
2.	Content author	Creator	Creator
3.	–	Subject	Subject
4.	Description	Description	Description
5.	–	Publisher	Contributor
6.	Website name	Contributors	
7.	archive_date, web year	Date	Date
8.	website_type	Type	Genre/Form
9.	Warc_file, file	Format	–
10.	Web_id, archive_id	Identifier	–
11.	File size	Source	[Extent]
12.	–	Language	[Source of Description]
13.	Foreign key	Relation	Relation
14.	–	Coverage	–
15.	Initial URL, website URL	Rights management (link)	URL
16.	–		Right

standard in web archiving repositories either ready shelves repository or self-development of web archiving project (Hazarika *et al.*, 2021; Ahammad, 2021; Evans, 2019).

Data elements that lacked in MFigures WA repository were identified as subject, language, coverage, right, publisher and contributor. Subject is important to the growth of ontology and taxonomy development in semantic web environments (Basyuk and Vasyliuk, 2021; Papasalouros and Chatzigiannakou, 2018). In Malaysia standard practices, language is used in websites mostly in Malay and English language. However, due to advancement of web emerging technology the language used in web content can easily be translated and changeable. Thus, creation of web content in multiple languages in Malaysia has become common to make it easily understandable by multiracial and multilingual people used while reading among Malaysian citizens. Coverage of metadata fields can reflect the broad and depth of website content scope in portraying the role figures. Metadata elements such as author, right, publisher and contributor can facilitate the knowledge discovery process and ownership.

Conclusions

The MFigures WA repository was designed to give an idea for database design and creation focus on preserving social and cultural web archive content collection. It also includes the assimilation of metadata standard mapping, which revealed that some other identification of metadata elements should be included in the next revision. Using ER Diagram, it is a clear view of each data element included in each table and the relationships; however, class diagrams also can be beneficial and assist on real-world objects' view. Novice web archivists, researchers and students will easily understand the data elements resided on ER diagram visualization. Each data element later on can be substituted to develop a higher level of information architecture on the web such as taxonomy on Malaysian social and cultural heritage. The MFigures data model can be adapted or transformed into real metadata modelling or UML class diagram for leveraging the physical view of web archive repository or implementation of web archive repository.

References

Ahammad, N. (2021), "Quality control (QC) of an institutional repository: a hands-on", *Collection and Curation*, Vol. 40 No. 4, pp. 145-152.

Amin, R. *et al.* (2012), "Transforming model to meta model for knowledge repository of Malay intangible culture heritage of Malaysia", *International Journal of Electrical and Computer Engineering*, Vol. 2 No. 2, p. 231.

Amran, N., Mohamed, H. and Bahry, F.D.S. (2018), "Developing human resource training management (HRTM) conceptual model using entity relationship diagram (ERD)", *International Journal of Academic Research in Business and Social Sciences*, Vol. 8 No. 12, pp. 1444-1459.

Ayadi, A., Rose, J., de Garidel-Thoron, C., Hendren, C., Wiesner, M.R. and Auffan, M. (2021), "MESOCOSM: a mesocosm database management system for environmental nanosafety", *NanoImpact*, Vol. 21, p. 100288.

Ayala, B.R. (2020), "Correspondence as the primary measure of quality for web archives: a grounded theory study", In *International Conference on Theory and Practice of Digital Libraries*, Springer, Cham, pp. 73-86.

Basyuk, T. and Vasyliuk, A. (2021), "Approach to a subject area ontology visualization system creating".

Brügger, N. and Laursen, D. (Eds) (2019), *The Historical Web and Digital Humanities: The Case of National Web Domains*, Routledge.

Coronel, C. and Morris, S. (2019), *Database Systems: Design, Implementation, and Management (13th ed.)*, Cengage Australia.

Di Pretoro, E. and Geeraert, F. (2019), "Behind the scenes of web archiving: metadata of harvested websites", *Archives et Bibliothèques de Belgique-Archief-en Bibliotheekwezen in België*, Vol. 106, pp. 63-74.

Dooley, J. and Bowers, K. (2018), "Descriptive metadata for web archiving: recommendations of the OCLC research library partnership web archiving metadata working group", OCLC Research.

Eichler, R., Giebler, C., Gröger, C., Schwarz, H. and Mitschang, B. (2020), "HANDLE-A generic metadata model for data Lakes", *International Conference on Big Data Analytics and Knowledge Discovery*, Springer, Cham, pp. 73-88.

Evans, E. (2019), "Semantic interoperability of moving image records in aggregating digital repositories: a comparative study of Dublin Core and PBCore descriptive metadata schemas.

Giannoulakis, S., Tsapatsoulis, N. and Grammalidis, N. (2018), "Metadata for intangible cultural heritage", *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP 2018)*, pp. 634-645.

Hazarika, H.J., Ravikumar, S. and Handique, A. (2021), "DSpace information retrieval system: a study using DICOM metadata standard", *Library Philosophy and Practice*, pp. 1-15.

Irfan, U., Mokhtar, S. and Abdul Mutalib, A.R. (2018), "A digital platform to (re)link heritage of Malaysia: co-producing information concerning ICH/Irfan Ullah, Mokhtar Saidin and Abdul Mutalib", *Idealogy*, Vol. 3 No. 2, pp. 29-36.

Khan, M. and Rahman, A.U. (2019), "A systematic approach towards web preservation", *Information Technology and Libraries*, Vol. 38 No. 1, pp. 71-90.

Lendák, I., Indig, B. and Palkó, G. (2020), "WARChain: blockchain-based validation of web archives", *International Workshop on Socio-Technical Aspects in Security and Trust*, Springer, *Cham*, pp. 121-134.

Momin, S.S. and Gaonkar, R.C. (2016), "Institutional repository – a gateway of global visibility: case study", *Research Dimensions*, Vol. 8 No. 2, pp. 103-110.

Musa, S.H., Sauti, N.S., Nasir, F.M., Abdullah, N.A. and Ahmad, Y. (2020), "The development of GIS-based digital archive for heritage building in Melaka world heritage site (UNESCO), Malaysia", *International Journal of Engineering Advanced Research*, Vol. 1 No. 3, pp. 1-16.

Németh, M. and Drótos, L. (2019), "A blended learning-based curriculum on web archiving in the national Széchényi library", *Digital Library Perspectives*.

Papasalouros, A. and Chatzigiannakou, M. (2018), "Semantic web and question generation: an overview of the state of the

art", *International Association for Development of the Information Society*.
Schafer, V. and Winters, J. (2021), "The values of web archives", *International Journal of Digital Humanities*, Vol. 2, pp. 129-144.
Sugimoto, S., Kiryakos, S., Wijesundara, C., Monika, W., Mihara, T. and Nagamori, M. (2018), "Metadata models for organizing digital archives on the web: metadata-centric projects at Tsukuba and lessons learned", *International Conference on Dublin Core and Metadata Applications*, pp. 95-105.
Talib, A.Z. and Osman, M.A. (2020), "Opportunities and challenges in enhancing access to metadata of cultural heritage collections: a survey", *Artificial Intelligence Review*, Vol. 53 No. 5, pp. 3621-3646.
Tan, Y.Y., Rafi, A., Musa, S.N. and Anuar, A.A.M. (2019), "Design and development of e-Warisan senibina portal: a web-based knowledge system for architectural virtual heritage data and heritage education", *Proceedings of the 2019 5th International Conference on Education and Training Technologies*, pp. 130-134.
Taylor, N. (2017), "Introduction to the special issue on web archiving", *Journal of Western Archives*, Zervas, M. Ktisis: Building an Open Access Institutional and Cultural Repository, Vol. 8 No. 2, p. 1.
Tomasi, F. (2018), "Modelling in the digital humanities: conceptual data models and knowledge organization in the cultural heritage domain", *Historical Social Research/ Historische Sozialforschung. Supplement*, Vol. 31, pp. 170-179.
UNESCO (2021), *Digital Heritage*, https://en.unesco.org/themes/information-preservation/digital-heritage.
Vlassenroot, E., Chambers, S., Lieber, S., Michel, A., Geeraert, F., Pranger, J., Birkholz, J. and Mechant, P. (2021), "Web-archiving and social media: an exploratory analysis", *International Journal of Digital Humanities*, Vol. 2, pp. 1-22.
Wijesundara, C. and Sugimoto, S. (2018), "Metadata model for organizing digital archives of tangible and intangible cultural heritage, and linking cultural heritage information in digital space", *LIBRES: Library and Information Science Research Electronic Journal*, Vol. 28 No. 2, pp. 58-80.

Further reading

Drótos, L. and Németh, M. (2018), "Web museum, web library, web archive the responsibility of public collections to preserve digital culture", in Petrovska, L., Īvāne-Kronberga, B. and Meldere, Z. (Eds), The Power of Reading: Proceedings of the XXVI Bobcatsss Symposium, Riga: The University of Latvia Press, Riga, Latvia, January 2018, pp. 124-126. Retrieved from http://bobcatsss2018.lu.lv/files/2018/08/BOBCATSSS_2018_TheProceedings.pdf
Golub, K., Milrad, M., Lundman, M. and Herault, R.C. (2019), "Increasing visibility of culture through online information services: the case of småland", *BOBCATSSS 2019: The 27th symposium*, 22-24 January 2019, *Osijek*, European Association for Library and Information Education and Research, pp. 458-470.

Corresponding author

Farrah Diana Saiful Bahry can be contacted at: farrahdiana@uitm.edu.my