E–B
Environment - Behaviour
Proceedings Journal

ITMAHuB2023

https://kppim.uitm.edu.my/research/conferences/itmahub-2023

**2nd International Conference on Information Science,
Technology, Management, Humanities, and Business**
Hotel UiTM, Shah Alam, Malaysia, 26-27 Sep. 2023
Organised by College of Computing, Informatics and Mathematics, UiTM, Malaysia

e-IPH
e-International
Publishing House Ltd.,
United Kingdom

Malaysian Homeland Warriors' Stories:
Database repository design and implementation

Farrah Diana Saiful Bahry, Farhah Nor Azam, Nur Athirah Md Ariffin, Noraizan Amran

School of Information Science, College of Computing, Informatics and Mathematics,
Universiti Teknologi MARA, Puncak Perdana Campus of UiTM Selangor, 40150, Malaysia

farrahdiana@uitm.edu.my, farhah501@gmail.com, 2020602556@student.uitm.edu.my, noraizan@uitm.edu.my
Tel:60192084038

Abstract

Increasing demand for information retrieval on the World Wide Web and the need for a database repository to efficiently manage web content. Aim to design a repository focused on Malaysian Homeland Fighters, ensuring the secure preservation and retrieval of legitimate information about Malaysian historians and fighters. Combination of the Web Archive Life Cycle (WALC) and Database Life Cycle (DBLC) methodologies used to identify relevant web content, design the repository, and develop a user-friendly interface. This initiative aims to preserve and authenticate historical web content, serving as a knowledge hub for Malaysians to connect sustainably with the past.

Keywords: web archive repository, web preservation, Malaysia Homeland Fighters, historical heritage

1.0 Introduction

Malaysia as a developing country is shifting way forward dashing into the Industrial Revolution 4.0 transformation with the advancement of Artificial intelligence (AI), the Internet of Things (IoT), and cloud computing technologies through digitisation (Hawkins, 2022; Ruest et al., 2021). Nevertheless, to improve the social well-being of Malaysian citizens, engaging the citizens with historical knowledge will make people become critical thinkers, argue with shreds of evidence to develop contextual understanding, and foresee a better future. Throughout history, people recognise their identity and grassroots of family ancestors and inherit their past ancestor spirits in leading and motivating current and future citizens as they navigate the complexities of modern life. The historical warriors or Malaysia homeland fighters are role models as good groundbreakers with full courage and big souls to protect the homeland. Thus, Malaysia's homeland stories in WebSphere must be preserved, representing in valid and legitimate source.

Malaysian historical warriors and fighters of the nineteenth century used well-thought-out battle strategies in conjunction with excellent fighting spirit and courage when confronting their enemies. Each of them has its own prodigious stories and contributions. Some historical warriors' and fighters' stories are well documented in the manuscripts or published material while some have already been transformed into digital heritage documents. However, the meagre effort made was insufficient to educate Malaysian citizens about the importance of their country's history and heritage. Preserving these historical warriors' stories is valuable for researchers, historians, and anyone to understand the past and build a stronger nation in the future.

Digital heritage documents were also included in web content information widely produced or reproduced on the World Wide Web (WWW). Web documents such as websites and social media content are the tangible historical heritage used to review, discuss, and share Malaysia's historical warriors and fighters. The information about the Malaysian historical warriors and fighters was found on the website or other web documents, which were integrated into digital object formats such as textual information, visual images and photos, audio-visuals, and video formats. However, the information content on the web still needs to be preserved, or it can be deleted due to web decay and loss of information content (Ben-David, 2019; Ka'ai-Mahuta, 2019). Thus, web preservation efforts must be cultivated and become a significant concern by the related parties.

1.1 Research Objectives and Scope
The web archiving research aims to preserve web content related to Malaysian historical warriors and homeland fighters who have become saviours in the past. Web content is a source of information for Malaysian citizen web users, especially students and researchers who need to use valid information, for instance.
1) To organise which web content of Malaysia historical warriors and homeland fighters is crucial to preserve.
2) To design a web archiving database repository for collecting Malaysian historical warriors and homeland fighters.

2.0 Literature Review
Archiving web content has become increasingly essential in preserving the cultural and historical heritage of nations, including Malaysia. In the context of Malaysia, where rich historical narratives intertwine with the struggles and sacrifices of homeland fighters, the need to archive web content related to this heritage is paramount. This literature review examines the importance of web archiving, challenges, and web preservation initiatives in securing Malaysia's homeland web content.

2.1 The Importance of Web Archive Content
The web archiving process normally is a systematic process that captures any web content updated by the web document, such as a website. Web archiving strategies can be categorised as site-centric or topical domain-centric (Khan & Rahman, 2019). Archiving the historical content on the World Wide Web is considered a specified or thematic archiving approach. Digital humanities linked data project had established social relationships between Jazz musicians viewed on visualisation tool, and Hawkins (2022) also claimed that facilitated by web archive service and linked data technologies able to sustain the digital cultural and historical heritage data on the web. Archiving indigenous digital content also becomes on-demand keeping all existing data either on the web, verbal or visual information in a secure knowledge information system or repository (Balogun and Kalusopa, 2022; Ayala, Hitchcock, and Sun, 2019; Ka'ai-Mahuta, 2019).

Web content related to Malaysia's history and culture, such as homeland heritage, historical wars, and events, should be captured and preserved. The Malaysia Ministry of Education has taken the initiative to make the history subject required to pass to all Sijil Pelajaran Malaysia (SPM) candidates since 2013. Ibrahim et al. (2021) also take one step further in designing an application for learning the homeland's history using Augmented Reality to attract students to learn history and make the students appreciate the homeland fighters. Within five years, there were also other initiatives such as publishing books and a complete study thesis related to Malaysia's homeland heritage (Affifudin, 2021; Hamzah, 2021, Affifudin, 2019). However, more than the scanty effort made is needed to make Malaysian citizens aware of their homeland's historical and heritage values and thus appreciate whatever has been done in the past.

2.2 Designing A Web Archive Repository
A repository is a place where items, whether in soft copy or hard copy form, are kept for safekeeping. The term repository is also used to describe stored materials, whether tangible or intangible. Web archiving, for example, captures and archives webpages for long-term access to assist in preserving web material using a crawler tool. Web-based and non-web-based archives can also be used to store web archives in addition to local files and local drives on the computer. The original URLs and linkages are kept while the web pages and accompanying metadata are gathered and stored in container files in a web-based archive. The non-web-based archiving method, on the other hand, extracts web content. The original URLs and linkages are kept while web pages and accompanying metadata are gathered and stored in container files in a web-based archive. Meanwhile, the non-web-based archive method pulls web documents from their hypertext environment and reorganises or converts them to PDF files for catalog-based access.

The database management system's (DBMS) capabilities and features have become more powerful in current database technologies updates. Most offered DBMSs support cloud-based environments, and some are even embedded with blockchain technology, which becomes more efficient, secure, resilient, and transparent (Hwang, Shon, & Park, 2020). Designing one's own database repository for storing web archive collection may reduce the issues of ownership, insecurity, and incompleteness that may be counted by using the proprietary software repository (Helmond and Van Der Vlist, 2021). Mkpojiogu (2020) described and reviewed six web-based archival system implementations that agree on the basic features needed, such as user login and management, data uploading, and searching (retrieval). The researchers also conform to the crucial features of web archival systems, which are multi-versioning of content management and metadata annotations for managing the evolution of web information content on the World Wide Web.

Malaysia Homeland Fighters (MYHF) database repository was designed to deposit web archive documents in WARC format and web archive content derived from web objects of the existing website. WARC file is an ISO-certified web archiving document that represents the whole description of the archive content with high authentication. Web objects extracted from archived websites must be

associated with metadata descriptions to reveal meaningful intrinsic values on each object. Metadata elements such as archive date, title, duration, file size, and the number of files archived were captured from the archive log and stored as technical metadata. However, descriptive metadata that semantically describes the intrinsic value of the web content needs to be curated by the curator or at least validated by the web owner or author.

3.0 Materials and Methods
The Database life cycle (DBLC) (Lirag et al. 2021; Haubt and Jalandoni, 2019) and Web Archiving Life-Cycle Model (WALCM) (Balogun and Kalusopa, 2022) attempt to bring together the technological and application aspects of the web archiving into a framework that will be a helpful reference either for a vast, small, or specified web archiving initiatives. Figure 1 shows the Web Archiving Life-Cycle Model (WALCM) which is made up of four essential components which are Policy, WALCM's Second Outer Circle, Metadata, and WALCM's Inner Circle. Nonetheless, depending on the web archiving technologies and resources available, the methodology for web archiving can vary and be adaptable for the implementation of specified web archive initiatives.
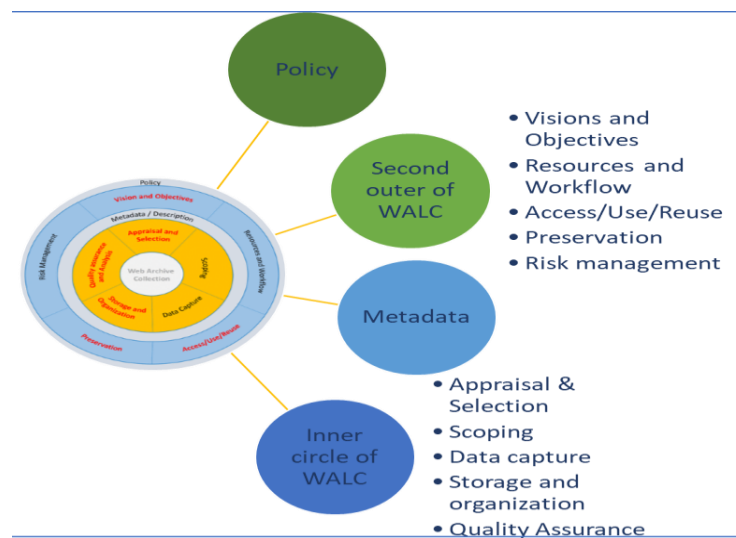

Fig. 1 Web archiving life cycle model (WALCM)

The second outer of WALCM is an initial insight related to the web archive initiatives which comprise visions and objectives, resources and Workflow, Access, Use & Reuse, and Preservation & Risk Management plan that should be thought off before proceeding to operationalise the web archiving process. The study is considered a small-range scale in the initial implementation, using the free crawler, and estimate to collect web archive content limit of 5GB of storage space for each warrior's stories (each MYHF). To prolong the lifespan of the web archive collection, a better web preservation strategy needs to be in place, proposing to the state library or any agency to continue the wheel.

In the inner circle of WALCM, which describes the day-to-day tasks involved in archiving the web activities, appraisal and selection, scoping, data capture, storage and organisation, and lastly, quality assurance takes place. All these are core steps in the web archive process. In this paper, the researcher focuses more on the first 4 of the inner circle of WALCM's components in exposing the Malaysia Homeland Fighter's web archive database repository development.

The appraisal and selection phase was initially started with brainstorming based on an individual case study of multiple Malaysia homeland warriors and fighters' stories extracted on the web. The main inclusion criteria are names of the Malaysian warriors or fighters followed by access to web documents. Primarily, 20 websites for each warrior were selected and listed in the web archiving seedlist. Each warrior's stories were then scoped and its content values identified. The brainstorming session occurred among the research teams, advisors, and validators. In the end, assisted by cloud-based repositories tools (XML & Builwith.com), information about website profiles and website content of each website URL in the seedlist was verified. The process is vital to ensure each website can crawl easily within an ample timeframe during the next stage.

During data capture, multiple crawler tools such as HTTRACK and Conifer were used for web archiving implementation due to their ease of use and free cost. Both tools produced different archive documents, such as Conifer producing the WARC file and HTTRACK producing the original web file and other related web documents. All web archive content can be preserved by storing it in the repository. Organisation of the web archive content can be designed by fulfilling the archive content preferences, such as via timeline, the structure of subject matter, or even categorical events or themes.

In order to confirm the accuracy and completeness of the web archive content gathered, the validation process has to be taken as a step of quality assurance. Each thoroughly archived website will be checked and rechecked for the web contents and also previewed back to the website archive via the link provided by HTTRACK. Using the Conifer tool, there is also a feature to replay the WARC file

that has already been archived. The inner circle of WALCM had taken about three and a half months to complete. This includes designing the database design for the MYHF repository for the best web archive content organisation.

4.0 Findings

:

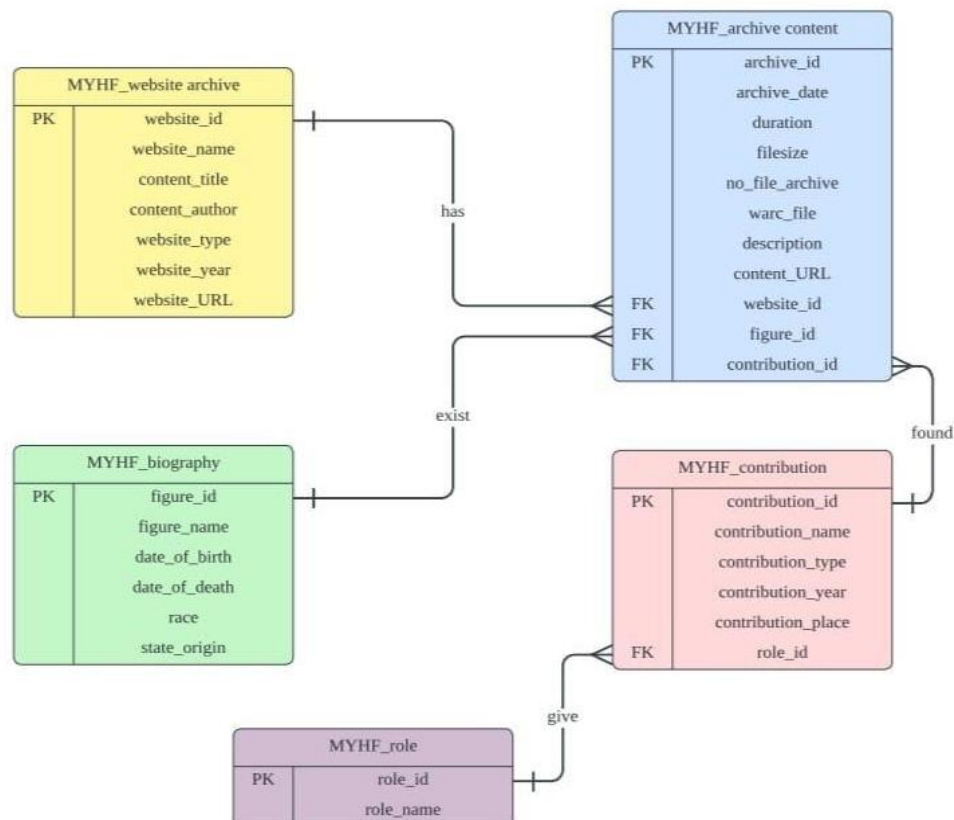*Fig. 2 Brainstorming output of Malaysia Homeland Warrior's Web Archive content*

Fig. 3 Malaysia Homeland Warrior Conceptual ER Diagram

The web archive database design was initially brainstormed based on an individual case study of multiple Malaysia homeland warriors and fighters' stories extracted on the web. Mainly, 20 websites for each warrior were selected and listed in the web archiving seedlist. Each warrior's stories were then scoped, and their content values were identified. The brainstorming session occurred among the research study teams, advisors, and validators. The output on the web content scoping is shown in Figure 2

The formal business rules for the MYHF database repository were then detailed as a proper guideline for implementing the database repository. Each website can have several web archive contents with different timestamped or content descriptions and URLs. Each web archive contains a story about warriors and their roles or achievements. The warrior's achievement in the battle event, victory place, year, and awards also had been extracted and captured.

Entity MYHF_website_profile captures all of a web page's information, while entity MYHF_archive content stores the multiple descriptions of the web archive processes; both are important as web preservation facts. Aside from that, the MYHF_biography, MYHF_role, and MYHF_contribution entities are important as either descriptive metadata or technical metadata. MYHF_biography entity consists of figure_id, figure_name, date_of_birth, date_of_death, race and state_origin. Meanwhile, MYHF_contribution becomes important for keeping the historical metadata element that consists of contribution_id, contribution_name, contribution_type, contribution_year, contribution_place, and role_id as referral fields to another table called MYHF_role that determine either the role of historian fighters or historian scholarly warriors. The eagle-eyes view of entities and relationships designed for Malaysia Homeland Fighter's database repository is depicted in the conceptual entity relationship diagram in Figure 3.

Consisting of 5 entities and each is related to one other. The entity-relationship (ER) diagram has been designed according to understanding and requirements as in the business rules (BR) below:

Table 1. Business Rules and entity's relationship descriptions

Business Rules	Entity's relationship descriptions
BR1: One website archive has many archive contents (1:M)	A website archive has one archive content (1:1) **or** a website archive has many archive contents (1:M). However, each archive content only belongs to one website archive (1:1).
BR2: One biography can exist in many archive contents (1:M)	Each biography of a fighter can exist in one archive content (1:1) **or** each biography of a fighter can exist in many archive contents (1:M). However, each archive content only stores about one biography of a fighter (1:1).
BR3: One contribution can be found in many archive contents (1:M)	A fighter contribution can be found in one archive content (1:1) **or** a fighter contribution can be found in many archive contents (1:M). However, each archive content only includes one fighter contribution (1:1).
BR4: One fighter role can give many contributions (1:M)	A fighter role can give one contribution (1:1) **or** a fighter role can give many contributions (1:M). However, each contribution is given by one fighter role only (1:1).

Fig. 4 Database structure of Malaysia Homeland Warrior

139

Figure 4 shows the database structure for the Malaysian Homeland Fighters Repository, which was implemented using a Free PHP-coded program called phpMyAdmin. PhpMyAdmin is designed to manage MySQL administration on the internet. Numerous MySQL and MariaDB operations are supported by phpMyAdmin. This user-friendly web interface has support for the majority of MySQL functionality, including the ability to build, copy, drop, rename, update, and browse the databases, tables, views, fields, and indexes, maintenance server, and execute, amend, and bookmark any SQL statement, including batch-queries (PhpMyAdmin, 2003).

Consist of 5 tables that aligned with the ER diagram, created to record and store the description of the websites that needed to be archived, every specific web archive procedure that was carried out for all linked websites, versions of the same websites being archived as well as information about Malaysian fighters scope of contribution and areas of skill or expertise. A complete description of the online archive process, including archive date and time, website owner, website content, and last changed date, must be documented to construct a web archive repository. The conceptual value of each web content format, whether visual, textual, or other media, must be adequately categorised or captured to ensure successful access and retrieval of web material in the future. In this situation, metadata significantly contributes to preserving variety across most internet platforms (Vlassenroot et al., 2022).
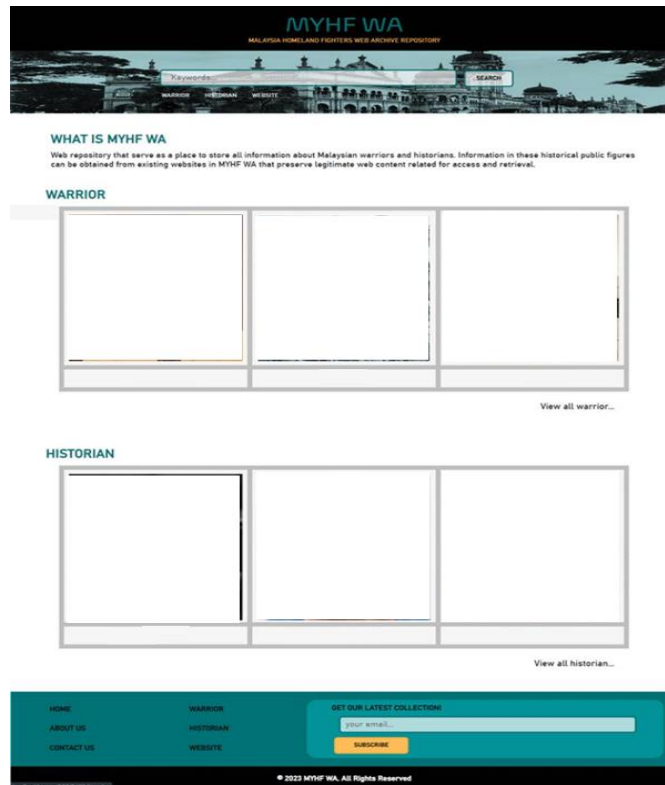

Fig. 5 Homepage of Malaysia Homeland Warrior

Figure 5 depicts the homepage's view of MYHF that enables web users to locate web archive content according to warrior or historian's name, years of events, or even the website's name. Homepage was embedded with web repository features such as a search option and a categorical web archive according to a list of website titles, fighters' names, or even historian contributions. Other information about us, the authorisation agency's contact information, and a subscribing option that allows web users to stay tuned with the latest update published by the MYHF repository. Malaysian Homeland Fighters Repository is not only targeted to deposit the web archive file and URL. However, it is an information discovery tool for Malaysian citizens to get information about Malaysian historians and warriors. The repository interface and further functionalities have been progressively implemented using PHP code. The web application of the MYHF is the second phase of the research study milestone, which is progressively in the development phase.

4.0 Discussions

Web preservation is crucial to ensure historical moments, events, and developments are documented and accessible. Historians, researchers, and anyone interested in understanding the evolution of society, technology, and culture may refer to valid web sources. In a similar initiative, the indigenous knowledge system (Balogun and Kalusopa, 2022) created long-term digital databases as inventories or registries to keep South Africa's traditional knowledge while (Ka'ai-Mahuta, 2019) focusing on collecting web archive content in repositories of 'truth' about Māori society in New Zealand. Martins and Rockembach (2020) expressed Harvard University's Web Archiving Initiative in collecting organisations and preserving a business history of companies, industries, individuals, and local movements, and Harvard Business School's institutional history, which total 69 collections, composed of the capture of 17,083 URLs

inside and outside the University's digital environment. All collections aimed to attract researchers, investors, and historians to research, teach, and preserve the University's institutional history.

No doubt, this web archive repositories or historical knowledge system can be a tool for delivering educational initiatives that are useful and designed to gain knowledge of society and raise public awareness about homeland warriors (Ibharim et al., 2021; Brata, Rai, and Seloka, 2021). The narrative of nationhood also nurtures the spirit of national integration between regions, tribes, cultures, religions, and beliefs through a series of histories as well as the struggles of our warriors while fighting against colonialism. Considering the advancement of web technologies nowadays, web preservation plays a vital role in consolidating prior knowledge. The substance of a website is given more weight on information content values when the archiving process goes through the proper web archiving process guidance such as appraisal and selection stage until organisation and storage to quality assurance (Vlassenroot et al., 2022; Ayala, Hitchcock, and Sun, 2019). On top of that, descriptive and technical metadata are also crucial in web archiving in giving more meaningful and retrievable archive content in the future (Formenton and Gracioso, 2023; Di Pretoro and Geeraert, 2019). The MYHF repository maintains the archival content of Malaysian warriors by embedding descriptive and technical metadata adapted partly from the Dublin Core metadata standard.

The Malaysian Homeland Fighters Repository was designed to inspire database design and development, focusing on preserving specific web content and facilitating users' access to legitimate information about Malaysian historians and warriors. Upon completing the research study, The MYHF repository will be proposed to be maintained by the Malaysia National Library or any library interested in handling this continuous effort to locate the web content related to Malaysia's homeland fighters and warriors and update in MYHF repository regularly.

5.0 Conclusions and Recommendations.

In an age where digital information is integral to our lives, web preservation plays a crucial role in preserving knowledge, culture, and history. It ensures that the digital footprint of our society remains accessible and usable for future generations and provides essential resources for research, education, and governance. Proper management and administration of the web archive content will ensure that the aim of web preservation will achieved. Besides that, the information contained in the ER Diagram representation is straightforward to interpret even for inexperienced web archivists, experts, and students, and easy to replicate. By using ER Diagrams, it provides a transparent perspective and interactions between every data element that is included in the tables. Additionally, it incorporates the metadata element which covers Malaysian warrior profiles and historical contributions that the subsequent revision should be taken for more excellent coverage. In the future, the study also recommends integrating more transnational metadata in web preservation efforts where interoperability and shared gateway can stimulate more new knowledge.

No doubt, some limitations have been found in the study such as a need for more discussion on the theoretical perspectives, including the framework development on web preservation toward preserving knowledge. From a practical point of view, researchers admit that the absence of vital fields on linked tables, such as subject, language, coverage, right, and contributor, has been observed and will be included in future improvement. On the contrary, with increasing information needs on the World Wide Web, more similar practices, such as implementing web archive collections and preservation, will be continuously established and improved for a broader range and more excellent coverage of digital archive content. Libraries and information agencies can play a significant role in enforcing web preservation through strategic planning and collaboration, such as establishing web archiving programs and integrating preservation into Library Services partnerships with key players in emerging technology providers. Another essential recommendation is that all those web archiving efforts and initiatives must reach relevant parties such as the content owner and even the society. Thus, more educational and awareness programs can be organised. Educating and developing the knowledge society is a long-term effort and should be continued.

Acknowledgements

References

Affifudin, A. H. (2019). Performing Silat in Malaysian diasporic communities in Australia: a cultural heritage analysis of identities, performance and self.

Affifudin, A. H. (2021). *Performing Silat in Australia: A Cultural Heritage Approach (UUM Press)*. UUM Press.

Ayala, B. R., Hitchcock, E., & Sun, J. (2019). Using image similarity metrics to measure visual quality in web archives. *Web Archiving & Digital Libraries*, 4.

Balogun, T., & Kalusopa, T. (2022). Web archiving of indigenous knowledge systems in South Africa. *Information Development*, *38*(4), 658-671.

Brata, I. B., Rai, I. B., & Seloka, I. B. (2021). National Heroes in the Indonesian Revolution and the Meaning for Young Generation. *International Journal of Social Science*, *1*(4), 407-414.

Ben-David, A. (2019). 2014 not found: a cross-platform approach to retrospective web archiving. *Internet Histories*, *3*(3-4), 316-342.

Formenton, D., & Gracioso, L. D. S. (2023). Metadata standards in web archiving technological resources for ensuring the digital preservation of archived websites. *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, *20*.

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

Hamzah, S. A. (2021). Branding the nation through Ahmad Nawab's "Malaysia Truly Asia". In *Made in Nusantara* (pp. 153-163). Routledge.

Haubt, R. A., & Jalandoni, A. (2019). Optimising the potential of research data through an integrated data management approach: considering research method, data life cycle, big data and linked data in an eResearch example in Australian rock art. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *4*, 77-82.

Hawkins, A. (2022). Archives, linked data and the digital humanities: increasing access to digitised and born-digital archives via the semantic web. *Archival Science*, *22*(3), 319-344.

Helmond, A., & van der Vlist, F. (2021). Platform and app histories: Assessing source availability in web archives and app repositories. In The Past Web: Exploring Web Archives (pp. 203-214). Cham: Springer International Publishing.

Hwang, H. C., Shon, J. G., & Park, J. S. (2020). Design of an Enhanced Web Archiving System for Preserving Content Integrity with Blockchain. *Electronics*, *9*(8), 1255.

Ibharim, N. A. N., Ramli, S. Z., Zahari, S. A., Edyanto, N. A. A., & Zawawi, M. A. A. (2021). Learning History Using Augmented Reality. *International Journal of Multimedia and Recent Innovation*, *3*(1), 1-10.

Ka'ai-Mahuta, R. (2019). Preserving Indigenous voices: Web archiving in Aotearoa/New Zealand. *IxD&A*, *41*.

Khan, M., & Rahman, A. U. (2019). A systematic approach towards web preservation. Information Technology and Libraries, 38(1), 71-90.

Lirag, J. R., Atole, R., Bello, L. C., Artiaga, R. J., & Briones II, S. (2021). The Pili knowledge bank: A comprehensive digital repository backend for Pili (Canarium Ovatum Engl.) information system. *International Journal of Applied Science and Engineering*, *18*(4), 1-11.

PhpMyAdmin. (2003). PhpMyAdmin. PhpMyAdmin Contributor. Retrieved July 17, 2022,
from https://www.phpmyadmin.net/

Martins, M. R., & Rockembach, M. (2020). Mapping of audiences for academic web archiving initiatives. *Intercom: Revista Brasileira de Ciências da Comunicação*, *43*, 71-88. https://doi.org/10.1590/1809-5844202014

Mkpojiogu, E. O., Akusu, G. E., Hussain, A., & Hashim, W. (2020). Implementation of a web-based data archival management system. *International Journal of Advanced Science and Technology*, *29*(6), 307-319.

Rhizome. (2022). Conifer. Retrieved July 17, 2022, from https://conifer.rhizome.org/

Ruest, N., Fritz, S., Deschamps, R., Lin, J., & Milligan, I. (2021). From archive to analysis: accessing web archives at scale through a cloud-based interface. *International Journal of Digital Humanities*, *2*(1), 5-24.

Vlassenroot, E., Chambers, S., Lieber, S., Michel, A., Geeraert, F., Pranger, J., ... & Mechant, P. (2021). Web-archiving and social media: an exploratory analysis. International Journal of Digital Humanities, 1-22.