

Original article

International Journal of Documentary Heritage
(Int J Docum Herit, IJODH) 2024 December,
Vol.1 No.1, Article 3
<https://doi.org/10.71278/IJODH.2024.1.1.3>

Received: August 27, 2024
Revised: November 20, 2024
Accepted: December 05, 2024
Published: December 30, 2024

Analysing the “World Significance” of Documentary Heritage: Quantitative Research on the Memory of the World Register

Yujue Wang^{1,2}, Ya Zhu^{1,2}

¹School of Information Management, Wuhan University, Wuhan, China

²Intelligent Computing Laboratory for Cultural Heritage, Wuhan University, Wuhan, China

Correspondence to Yujue Wang, E-mail: Yujue.wang@whu.edu.cn

Abstract

Background: UNESCO established the Memory of the World (MoW) Programme in 1992 to preserve the collective memory of humanity and the globally at-risk documentary heritage. ‘World Significance’ is one of the most crucial criteria in reviewing nomination forms for the MoW Register. However, research on the MoW Register is limited, and focused studies on the ‘World Significance’ criterion are scarce, making project submission complicated. **Purpose:** This paper’s purpose is to analyse the “World Significance” criterion of the MoW from the perspective of nomination forms. Thereby raising the world’s recognition of documentary heritage’s world significance and assist future nominators to effectively highlight how their nominated heritage meets this criterion. **Methodology:** This paper conducted a novel quantitative investigation into the “World Significance” section of 390 nomination forms of inscriptions on the MoW Register, using Latent Dirichlet Allocation as a thematic clustering method. **Findings:** This paper extracted that “Time”, “Space”, “Form”, and “Value” are the main dimensions of the “World Significance” description, and “War and Politics”, “Culture and Exchange of Civilizations”, “Arts and Humanities and Social Sciences”, “Minority Groups and Human Rights”, and “Oriental Medicine and Society” are the important themes in the description of “World Significance”. **Value:** This paper innovatively applied thematic clustering to analyse the MoW field where qualitative research is more prevalent. It addressed the problem of insufficient understanding of the value connotation of the existing key accreditation criteria, and selects the “World Significance” criterion for research to enrich the value connotation of this criterion.

Keywords

Memory of the World Programme, analysis of nomination forms, Latent Dirichlet Allocation, world significance, quantitative research

Introduction

The Memory of the World (MoW) Programme was launched by the United Nations Educational, Scientific, and Cultural Organization (UNESCO) in 1992 with the objectives of safeguarding documentary heritage, facilitating access, and disseminating documentary heritage as well as raising public awareness of its significance and the need to preserve it (UNESCO, 2021b). And the General Guidelines of the Memory of the World Programme is the programmatic document of the MoW Programme. Beginning in 1996, the MoW Register has been reviewed every two years, with nomination forms received every even year and results announced every odd year. However, the International Advisory Committee (IAC) of the MoW initiated a comprehensive self-examination of how the programme was administered in 2015 because of Japan’s sceptical attitude toward the IAC’s “transparency and fairness” and authority (Suh,

OPEN ACCESS

pISSN : 3058-9428
eISSN : 3058-9061

© 2024 International Centre
for Documentary Heritage
under the auspices of UNESCO.



This is an Open Access article which is freely available under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND) (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

2020). After the comprehensive review, the General Guidelines was revised and the “Incidental Process” was added to it so as to deal with issues that arise when a nomination is contested. Moreover, the nomination process for the MoW Register was modified allowing nominations to be submitted by States Parties only through their National Commission for UNESCO (if not, through the relevant government body in charge of relations with UNESCO) (UNESCO, 2021b), with governments having increasing influence over the MoW Programme. As of 2023, the MoW Register has launched 12 review cycles with a total of 496 nominations, including documentary heritage, such as Nicolaus Copernicus’ Masterpiece “De Revolutionibus Libri Sex” and the Churchill Papers. This preliminary construction of a globally linked and mutually assisted documentary heritage protection system can jointly guard against the collective amnesia of the world caused by memory’s fragility (Harvey, 2003).

Although 2022 marked the 30th anniversary of the MoW Programme, studies on the programme itself were still scarce. Previous research has focused on the development of the programme, including the current issues and achievements in practice across multiple countries (for example, Harvey, 2007), spatial and temporal differences in national and regional participation in MoW registers (Dippon & Helferich, 2023), as well as the preservation status and strategies for the nominated documentary heritage (Roger, 2019; Kim et al., 2021a). Moreover, the scope of existing studies on the MoW Programme is generally limited, and the implemented research techniques are largely qualitative, homogeneous, and lacking the application of quantitative research methods. While the Latent Dirichlet Allocation (LDA) model has not been directly applied to the study of MoW Register nomination forms, it has been successfully used to cluster and analyse themes in tourist reviews of World Natural Heritage Sites (Kim et al., 2019; Kim et al., 2021b), demonstrating the feasibility of applying LDA to heritage research.

In addition, the most important selection criterion for the MoW Register is whether the submitted heritage has outstanding “World Significance” and common value, which should be stated clearly in its nomination form. Nevertheless, UNESCO has not yet defined the concept of “world significance” in its General Guidelines, which considers documentary heritage to only be of world significance if it meets any of the primary criteria of “Historical significance”, “Form and style significance” and “Social, community or spiritual significance” (UNESCO, 2021a). Compared with the 1995 General Guidelines, the “Historical significance” covers several criteria such as initial “Influence”, “Time”, “Place” and “Subject /Theme”. The “ Social, community or spiritual significance “ pointing to social value covers the original indicators such as “People” and “Social Value”, and the “ Form and style significance “ standard has always been the key point emphasized by MoW (Foster et al, 1995). This shows that the connotation of the “world significance” value of documentary heritage is constantly enriched and improved with the practice of the project. The original seven indicators have been summarized into three, and the meaning and connotation of each standard is richer, and the value orientation of assessment is clearer. This will be conducive to the claimants to make a clearer description of the world significance and value of documentary heritage. At the same time, through the evolution,

we can find that the historical value and social value have gradually become the focus of determining the meaning of the documentary heritage, and the formal value still plays an important role from beginning to end. But we still don't have a clear definition of what the “world significance” standard means, and further critical reflection on the connotation of criteria is an urgent problem to be solved in current academic research (Jordan, 2020).

Accordingly, this research uses novel technical tools, such as the LDA model, to quantitatively analyse the “World Significance” component of the 390 nomination forms collected from the MoW Programme website. The following questions are explored: (1) What dimensions are used for describing the “World Significance” of world-class documentary heritage? (2) Which themes are the focus of attention in assessing international documentary heritage? In addition to defining the value of “World Significance” and enhancing the world's recognition of “World Significance” as the core concept of UNESCO's documentary heritage conservation, the current status of international documentary heritage is presented and the gaps in the development of documentary heritage are examined for different themes. Moreover, this study can assist countries select objects for nomination which are more in line with the “world significance” criterion and completing nomination forms to better demonstrate the characteristics that meet the needs of the MoW Programme.

Research Framework

This study presents quantitative and qualitative analyses on the “World Significance” section in the nomination form of documentary heritage as inscribed in the MoW Register. The implemented methodology includes word frequency statistics as well as LDA topic clustering. The pre-processing, modelling, and visualisation of the data were implemented in the python language, and the research framework is shown in Figure 1.

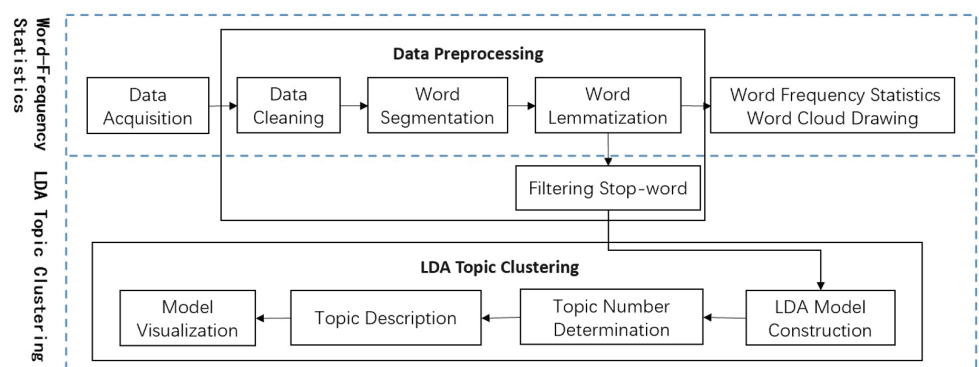


Figure 1. Research framework

Data Acquisition and Pre-processing

We selected the nomination forms from 1997 to 2017 as the data source. The nomination forms of documentary heritage inscribed in the MoW Register between 1997 and 2017 were obtained from the official website of the MoW Programme(<https://en.unesco.org/programme/mow>).

After eliminating nomination forms with incorrect data, those in French, and those without a statement of “World Significance”, 390 nomination forms were retrieved. The text in the “World Significance” section of each nomination form was extracted, and an experimental dataset was constructed, with 884,049 words in total.

After data cleaning and word segmentation, the text was transformed from sentences to words, eliminating invalid information, such as numbers and characters. Subsequently, the root words were extracted by lemmatisation to make the singular and plural forms as well as tenses consistent. Finally, a stop-word list containing 634 words was developed based on the characteristics of the MoW Register nomination forms. This list contains technical terms, such as “archive”, “document”, “world”, and other high-frequency words, which were then filtered according to the word list.

Word-Frequency Statistics

Following the word lemmatisation step, the frequencies of the words in the “World Significance” section of the nomination forms were calculated and sorted in descending order. Next, Python was used to create a visual word cloud to synthesise and analyse the hotspot issues reflected by high-frequency words.

LDA Topic Clustering

In order to further study the hot topic of “world significance” and reduce the influence of the length of the nomination forms on the word frequency clustering results this study primarily adopted the LDA model to achieve text clustering and potential topic exploration. The LDA model is essentially a three-level hierarchical Bayesian generative probability model with document, topic, and word as the three levels (Blei et al., 2003). It can find potential and hidden information from a text set by maximising the probability of word co-occurrence under a predetermined number of topics, achieving the goals of mining and clustering hidden topics in text.

LDA topic clustering consists of three main steps: model construction, topic number determination, and model visualisation. First, the pre-processed words were vectorised to construct a document-word matrix, and the implicit topic vector was extracted from the unstructured raw text to complete the construction of the LDA topic model through parameter tuning. Second, perplexity was used as an indicator to determine the number of topics, and a debugging range of k was initially set from 1 to 20. We established through model construction and extensive experimentation that when $k = 12$, perplexity was minimum and the model performs best; therefore, the final LDA topic model was constructed with $k = 12$. Finally, a web-based interactive topic visualisation tool was used to visualise the interactive LDA of the best models, resulting in a visual representation of the importance of each topic in the corpus and the similarity between all topics (Sievert & Shirley, 2014).

Results

Word-Frequency Statistics

The top 30 high-frequency words, as listed in Table 1, were obtained from the word frequency statistics of the processed nomination forms. Word frequency can measure the degree of repetition of a word in the text, and the importance of a word increases proportionally with the number of times it appears in the document. From the results of the word frequency statistics, we can get a glimpse of the focus of the “World Significance” section in the MoW Register. The word “world” appears most frequently with 750 occurrences. It appears nearly twice as often as the second-ranked word “collection” that corresponds to the main point of the “World Significance” section, which is stating the worldwide impact of cultural heritage.

Based on the obtained word frequency statistics, a visual word cloud map was built using the top 200 words in terms of word frequency to further reveal high-frequency hotspot issues. The resulting word cloud diagram is shown in Figure 2 in which a larger font size indicates a higher word frequency for that word.

A comprehensive analysis of the word frequency statistics and the visual word cloud map shows that the high-frequency words can be divided into four dimensions describing time, space, form, and value, as summarised in Table 2.

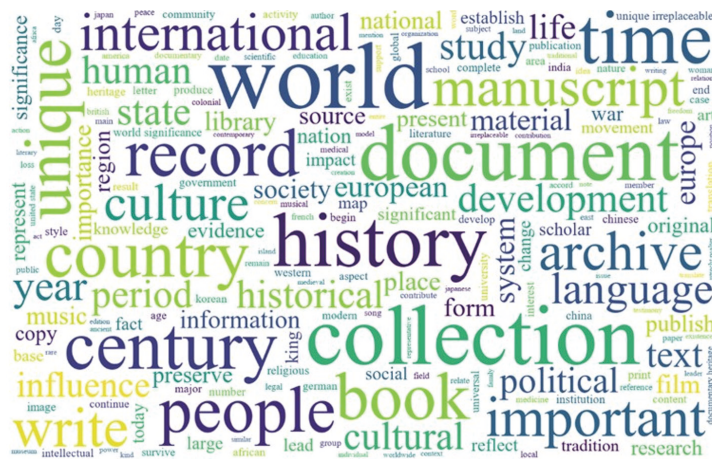


Figure 2. Word cloud for the ‘World Significance’ section of nomination forms

Table 1. High-frequency characteristic words from the ‘World Significance’ section (Top 30)

No.	Characteristic Words	Frequency	No.	Characteristic Words	Frequency	No.	Characteristic Words	Frequency
1	world	750	11	archive	266	21	significance	183
2	collection	420	12	international	265	22	language	183
3	century	398	13	record	263	23	political	175
4	document	391	14	manuscript	243	24	human	173
5	history	387	15	write	237	25	period	171
6	unique	347	16	state	233	26	development	169
7	time	309	17	important	224	27	heritage	169
8	country	281	18	cultural	216	28	historical	169
9	people	270	19	culture	190	29	national	163
10	book	267	20	year	189	30	influence	159

Table 2. Dimensional analysis of ‘World significance’ section

Dimension		High-frequency words (Word frequency)	
Time		century (398), time (309), year (189), period (171)	
Space	Location	World Level	world (750), international (265), global (67)
		Regional Level	europe (134), european (133), american (107), american (107), region (98), asia (96), africa (77), african (73), area (70), latin (52)
		National Level	country (281), national (163), nation (120), america (105), american (107), china (93), chinese (84), japan (77), french (71), india (66), korean (61), korea (57), india (56), german (59)
		Local Level	state (233)
Institution		archive (266), library (153), museum (52)	
Form	Emphasis on originality of record	document (391), archive (266), record (263), manuscript (243), documentary (110)	
	Emphasis on value of form and style	collection (420), book (267), art (86), psalter (68), song (55)	
Value	History (556)	history (387), historical (169)	
	Culture (406)	cultural (216), culture (190)	
	Politics (175)	political (175)	

Theme Clustering

After LDA training, 12 themes were extracted from the “World Significance” section of the MoW nomination forms. The results of the clustering and distribution of the top ten most relevant keywords under each theme are listed in Table 3. Figure 3 shows the LDA visual interaction diagram, where some circles representing topics overlap. It proves a high degree of similarity among the topics, due to the large differences in the length of each nomination form. Therefore, the existing 12 themes were clustered again.

The final clustering yielded five themes: “War and Politics”, “Culture and Civilizational Exchange”, “Arts and Humanities and Social Sciences”, “Minority groups and Human Rights”, and “Oriental Medicine and Society”. The specific thematic distribution with the top ten terms of relevance under each theme is listed in Table 3.

Table 3. LDA topic description

Topic	Topic Description	Lexical Item (Listed in the descending order of relevance)
Topic 1	War and Politics	shakespeare war collection polish film political union year history poland
Topic 12	Culture and Civilizational Exchange	map piri ¹ macao mission silk real saiva ² marti siddhanta ³ cartographic
Topic 8	Arts and Humanities and Social	psalter utrecht ⁴ psalm manuscript carolingian newton byzantine fuzuli ⁵ period david ⁶
Topic 9	Sciences	sa’ di ⁷ gulistān ⁸ conservation language wittgenstein ⁹ translation marx borobudur ¹⁰ masaalik ¹¹ century
Topic 11		bach ¹² art galigo ¹³ nezami ¹⁴ minor mass manuscript martial painting khamseh ¹⁵
Topic 3	Minority Groups and Human Rights	court international dhakhīra ¹⁶ central mexican american amparo ¹⁷ remedy labour jacob ¹⁸
Topic 4		constantine ¹⁹ cricket tripitaka ²⁰ lord collection mcluhan ²¹ papiamentu ²² koreana book print
Topic 5		diponegoro ²³ colonial javanese philippine university jazz africa country african leuven ²⁴
Topic 6		journal record farquharson ²⁵ imperial society broadcast speech family write bogomil ²⁶
Topic 7		indigenous collection melaka plantation people manuscript baroque region massage thai
Topic 2	Oriental Medicine and Society	music jewish ilseongnok ²⁷ collection joseon people film culture map song
Topic 10		uigwe ²⁸ compendium ²⁹ iici ³⁰ ritual bogam ³¹ disaster book confucian medicine tsunami ³²

Notes to Table 3 at the end of this article.



Figure 3. LDA visual interaction diagram

Discussion

The analysis of the 390 nomination forms from the MoW Register using the LDA method shows that “Time”, “Space”, “Form”, and “Value” are the main dimensions of the “World Significance” description. Moreover, the important themes in the description of “World Significance” are: “War and Politics”, “Culture and Exchange of Civilizations”, “Arts and Humanities and Social sciences”, “Minority Groups and Human Rights”, and “Oriental Medicine and Society”. The absolute majority of the nomination forms describe the global nature of nominations in terms of the aforementioned “form” and “theme”.

Dimensions of “World Significance”

Time: the creation and duration of the impact of documentary heritage

The terms “century”, “time”, “year”, and “period” occur a combined total of 1067 times in the top 30 high-frequency words. The high occurrence of time-related terms is linked to the “Authenticity” selecting criterion, which stipulates that the submitted heritage must be “real, true or genuine and not corrupted from the original” (UNESCO, 2021a). The beginning and evolution of the documentary heritage can be traced owing to its temporal definition and description, considerably enhancing its validity and dependability. This also echoes with the “Historical Significance” aspect of the “World Significance” criterion, which pays attention to the “Time” of documentary heritage. Additionally, the jury is helped in examining the facts and reaching a verdict on whether the documentary heritage has had a long-term influence on a specific cultural field in the world.

Space: the location, sphere of influence, and conservation institution

Based on word frequency statistics, space-related terms may be divided into two categories: place and organization.

The Memory of the World Register Companion reports that the emphasis on articulating “World Significance” lies in elaborating the magnitude of the geographical influence of documentary heritage (UNESCO, 2012). Consequently, nomination forms contain detailed mentions of locations that are either associated with a historical event that gave rise to the documentary heritage or where history and regional culture have been profoundly impacted by the documentary heritage. Based on the magnitude of the geographic area of influence, space may be classified into four levels: “world”, “regional”, “national”, and “local”.

At the world level, “world” is the most frequently used word, and the frequent occurrence of world terms, such as “international” and “global”, indicates that the MoW Register favours documentary heritage with a significant impact on the entire world.

At the regional level, the word frequency for Europe is larger than Asia, which in turn is larger than that for Africa (since “America” can refer to a continent or a country, it is not included in the rankings). These frequencies are broadly consistent with the ranking of documentary heritage regions in terms of the number of inscriptions, with “Europe and America” higher than “Asia and the Pacific”, “Latin America and the Caribbean”, “Africa”, and “Arab Region” (Bu & Wang, 2021). This demonstrates that the “World Significance” section also focuses on elaborating repercussions on an intercontinental and regional scale.

At the national level, the frequency of the national vocabulary is somewhat indicative of national nominations, but it cannot be directly correlated with the order of nations in terms of the number of nominations. This dimension is heavily affected by the length of the nomination form, the richness of the supporting material, the extent to which national contextual information is omitted, and the terminological and logical conventions of each country.

At the local level, terms like “province” and “city”, aside from “state”, are used less frequently. The primary purpose of these terminologies is to convey background knowledge about the heritage and its growth, which is unaccounted for in most countries, leading to their low frequencies.

Terms related to conservation institutions, such as “archive”, “library”, and “museum”, are the main places where documentary heritage is kept. The *General Guidelines* include “8.3.6.1.2 Condition” as part of “8.3.6 World significance: Comparative criteria”, considering that “the owners and custodians of inscribed documentary heritage are encouraged to publicize their status” (UNESCO, 2021a). Some nomination forms describe the state of conservation through the custodian as the environmental elements and management techniques offered by the conservation institution for documentary material are significant factors in determining the conservation status of documentary heritage.

Form: the “world significance” value carried by the physical nature of documentary heritage

“Form and style significance” (8.3.5.1.2) is one of the three sub-criteria of the primary criteria of “World Significance”. As emphasised by *the Memory of the World Register Companion*, the ideas of ‘the importance of the carrier as well as the content matters’ and “the artefact value of the carrier also has to be appreciated” (UNESCO, 2021b) are important. The carrier itself has its special value in addition to its function as a mean of conveying information about the content.

The terms used in nomination forms concerning the carrier (i.e., form) of the documentary heritage can be divided into two categories. The first category emphasises the document’s original recording character, such as “document”, “archive”, “record”, and “manuscript”. It focuses on the fact that the documentary heritage, as a record of historical events through the form of a vehicle, satisfies the criteria for authenticity and is authentic, complete, and undamaged. The second category concentrates on the artistic and cultural worth that the documentary heritage form offers to humanity by highlighting the art form and the value it carries through words such as “art”, “song”, “book”, “collection”, and “psalter”.

Furthermore, the unique carrier specificity of the documentary heritage has outstanding worldwide and memory values. In the “World Significance” section, nominators should highlight the documentary heritage’s irreproducible aesthetic and technological allure or argue that it represents the rarity of a type that is vanishing or has already vanished. Moreover, nominators should investigate the exceptional technical, artistic, and historical value of the documentary heritage.

Value: historical, cultural, political values and “World Significance” value

The main values of documentary heritage are historical, cultural, and political dimensions. According to the word frequency statistics, the use of history-related words such as “history” and “historical” was greater than the use of culture-related words such as “culture” and “cultural”, and more than the use of politic-related words such as “political”.

The MoW aims to preserve documentary heritage, which is the product of the coalescence of memory and culture over the course of human history. In the section on “World Significance”, history is highlighted to portray the origins and duration of influence, with culture emerging from the influence of documentary heritage. History is described more often than culture as historical origins are common to all documentary heritage objects; however, the impact of heritage is primarily but not exclusively cultural.

The political aspect of documentary heritage has received the least attention of the three as the MoW insists on its goal of safeguarding the collective documentary heritage of humanity and adheres to the criteria of depoliticising the documentary heritage selection process. To prevent international political disputes, the nominators also try to avoid highlighting political values when filling out their nomination forms. However, political values are always present as the discussion of historical events may be interspersed with a brief description of their political context.

Critical Topics in “World Significance” Section

War and Politics

For the MoW, “War and Politics” is a crucial theme. Although the MoW has always maintained a neutral political stance, it has always had an inclusive and open attitude towards the documentary heritage of war and politics, seeking to present a true and complete memory of the world.

Most documentary heritage preserved in the MoW Register reflects modern history, where the transition from a feudal society to a modern society is inevitably accompanied by the pain of political change and even bloody conflict. A great deal of the documentary heritage submitted by Poland has a direct connection to politics and warfare, such as *Jürgen Stroop’s Report*¹, *Documents of Polish Radio Intelligence from the Period of the Battle of Warsaw in 1920*, and *the Act of the Union of Lublin document*. The twists and turns of Poland’s fate reflect the changing political situation of the world at large.

Human civilisation has progressed and developed through constant contradiction and conflict. Considering the world as a whole, changes in the political ideologies of any country and wars between countries, profoundly impact the course of world development. One of the reasons for the documentary heritage of “War and Politics” becoming the focus of the MoW is that the history of war is an important component of human history.

Culture and Civilizational Exchange

The MoW’s principal goal is “to increase awareness worldwide of the existence and significance of documentary heritage and thereby foster dialogue and mutual understanding between people and cultures” (UNESCO, 2021a). The selection of global documentary heritage provides an invaluable forum for polite dialogue, exchange, and recognition among civilizations, further enhancing the understanding and appreciation of the value of cultural diversity between different societies and groups (UNESCO, 2011).

Compared to other critical topics, “Culture and Civilizational Exchange” is more narrowly focused; nevertheless, it has worldwide significance for tracing the history of civilizational exchanges beyond geographical barriers. The documentary heritage related to cultural and civilizational exchanges can be broadly divided into two types. The first type is the visual record of civilizational exchanges in the form of letters and archives, such as the *Official Records of Macao During the Qing Dynasty (1693–1886)*, which reflects Macao’s important role as a hub for China’s foreign trade and cultural exchanges through official correspondence between China and Portugal during the period of Portuguese administration. Another type is the indirect reflection of world connectivity through the medium of maps, as in the case of the *Piri Reis World Map (1513)*, which precisely depicts the coastline of the New World in the form of a circular map and first transmitted cultural and biological information from South America to the European continent.

¹ Nazi SS Major General, who was a German commander, suppressed the Warsaw Partition Uprising and edited the Jürgen Stroop’s Report.

Arts and Humanities and Social Sciences

The MoW’s focus area of “Arts and Humanities and Social Sciences” encompasses a wide range of subjects, including philosophy, science, ideas, and religion. As art is a means of cultural expression and communication as well as an essential tool for exchange between civilisations, the MoW has consciously included different forms of artistic expression in its identification of documentary heritage (Bu & Wang, 2021). Moreover, advances in the humanities and social sciences show a better understanding of the world and the development of human thought; therefore, humanities and social sciences are of equal world significance.

The archives of well-known individuals from various nations who made remarkable contributions to the arts, humanities, and social sciences at various times are a significant component of this theme. Te Baroque (17th–18th century) German composer Bach, the renowned English mathematician and physicist Newton, the 19th-century German thinker Karl Marx, the 20th-century British philosopher Ludwig Wittgenstein, and others are all included in the MoW Register.

Religion is a product of human spirituality and a significant component of traditional human civilization. Consequently, this topic also includes numerous heritage objects with overtly religious values, encompassing Christianity (Catholicism), Buddhism, Islam, Hinduism (Siddhanta Saivism), and other major global religions. Accordingly, religious heritage profoundly reflects regional customs and ideological states, as well as having its own unique cultural and social value.

Minority Groups and Human Rights

The topic of “Minority Groups and Human Rights” contains the words indigenous, women, black, slave, and colonial. These groups have consistently fought for their legal rights and defended their human rights despite having a lower social standing and a worse state of existence. As the protector of collective memory, the MoW Programme must give due consideration and respect to the heritage of all people, particularly the minorities and the more vulnerable members of society (Springer, 2020).

The documentary heritage of slave trade and colonial plunder consists of two aspects. One aspect is the exposure of the exploitative and oppressive lives suffered by the enslaved in the context of slavery and colonial rules, such as the *Farquharson’s Journal*, which documented the relationship between the enslaver and the enslaved from the perspective of plantation owners. The other chronicles the history and contributions of notable individuals that fought against racial discrimination and eased race relations, such as the Javanese nobleman Diponegoro, who is remembered for leading the Indonesian people of Java in their resistance against Dutch colonial oppression, and the *First Catechism Written in Papiamentu Language*, the first official political recognition of the existence of the “Papiamentu” language as a milestone in breaking down linguistic prejudice and racial discrimination in indigenous languages around the world.

The history of women's awakening and struggle is also remembered in documentary heritage and is a significant part of the world's memory as in the case of Aletta H. Jacobs, a doctor who fought for women's rights through essays and writings that form the foundation of the archives of the international women's movement. *The 1893 Women's Suffrage Petition* saw the birth of the world's first self-governing country where women won the right to vote.

Oriental Medicine and Society

The topic of "Oriental Medicine and Society" focuses on self-contained? Oriental medicine and reflects the overall development of Oriental societies. The Oriental societies here refers to the East Asian cultural circle, which refers to a group of countries and regions influenced by Chinese culture in history. It has become a critical topic for the MoW and a specific calling card for Oriental countries to present themselves to the world based on their distinct medical systems and cultural legacies that set this theme apart from the other themes.

The ancient East's culture is quite alluring to the West. Eastern nations have been actively submitting nominations for the MoW from its inception, enthusiastically promoting their exceptional documentary heritage and dedicating themselves to presenting Eastern culture to the rest of the world. Among the keywords under this theme, an abundance of references to Korea can be discovered. This is primarily because Korea has the greatest number of documentary heritage objects inscribed in the MoW Register in Asia, and the nomination forms are frequently lengthy. Moreover, the themes tend to centre on its history and culture, which is a good way to showcase Eastern characteristics to a wider audience.

The current inclusion of documentary heritage related to Oriental societies suffers from a lack of comprehensiveness and scale. Nowadays, the MoW project is dominated by a "western" viewpoint, which broadly influenced the ways of thinking and acting historically, as well as continuing to influence the interaction of ideas and knowledge today (Edmondson, 2020). Furthermore, the documentary heritage of Eastern societies is underrepresented in the overall world heritage system. In addition, Eastern countries have yet to form a comprehensive system of evidence and development of Eastern societies.

Conclusions

Based on word frequency statistics and thematic clustering in the "World Significance" section of the MoW nomination forms, the main findings of this study are as follows.

First, the "World Significance" section of a documentary heritage inscribed in the MoW Register is primarily articulated in terms of its "Time", "Space", "Form", and "Value". This conclusion may provide direction for the "World Significance" description of documentary heritage for future nominations to the MoW Register. For describing documentary heritage, temporal world representativeness, such as representing a particular historical period in the development of the world or witnessing a major historical change, can be a focus. In terms of space, the significance of documentary heritage must be emphasised in terms of its impact on

national, regional, local, and continental levels while adequately reflecting its value and sphere of influence. Additionally, the description of the institution where the documentary heritage is kept should be added, which proves that the documentary heritage is being effectively protected and its development is sustainable. Moreover, if the submitted documentary heritage possesses special value in terms of the carrier or material, it must be fully specified.

Second, based on the contents of the documentary heritage included in the MoW Register, “War and Politics”, “Culture and Civilizational Exchange”, “Arts and Humanities and Social Sciences”, “Minority Groups and Human Rights”, and “Oriental Medicine and Society” are important thematic topics. Even though the countries and positions of nominations are different, these five thematic topics are primarily represented. On the one hand, these topics reflect UNESCO’s “priorities” in the consideration of “having world significance”; on the other hand, they also provide a certain direction for future documentary heritage nominated for the MoW Register.

Third, the MoW maintains a balance between global “glorious memory” and “traumatic memory”. The MoW covers not only the glorious memories of the progress of global civilization, such as the accomplishments in arts, humanities, and social sciences, as well as the record of cultural and civilizational exchange and integration, but also the traumatic memories of wars, massacres, and the struggles of oppressed groups. Consequently, the MoW helps to some extent preserve “traumatic memories” in the global memory. In particular, since the World Heritage Programme has urged the heritage “associated with recent conflicts and other negative and divisive memories” to register for the MoW Register (World Heritage Convention, 2021), it may end up being the only international heritage list to record global trauma memory.

Fourth, the documentary heritage included in the MoW Register is characterised by regional “unevenness” and “western centrism”; Europe and North America have been nominated for 230 documentary heritage items, more than half of the total, followed by the Asia-Pacific region and then the Latin America and Caribbean region with 109 and 69 documentary heritage items, respectively. Moreover, the African and Arab regions are far behind with only 23 and 13 documentary heritage items, respectively. Additionally, according to the topics and contents submitted, the topics of “War and Politics”, “Culture and Civilizational Exchange”, and “Arts and Humanities and Social Sciences” are particularly focused on the Western perspective, reflecting a certain amount of “Western centrism” in the evaluation of the MoW Register as well as the neglect of other countries in the nomination of critical topics.

In conclusion, the addition of novel technological instruments offers a new perspective on the current state of the MoW’s preservation. This study provides directions for further developing and improving the MoW Register. Nowadays, the MoW Register primarily focuses on five critical topics, including the glorious and traumatic memories of human history, and finally constructs the “defences of peace”² by preserving historical accuracy. This working principle conforms with

² This is an excerpt from the Preamble to the UNESCO Constitution, as adopted on 16 November 1945 and amended at subsequent General Conferences.

UNESCO's concept of "Building peace in the minds of people"³. However, challenges still exist, such as regional imbalance, propensity for the "Western viewpoint", and neglect of nomination applications in some countries, which remains to be explored in the future.

Limitations and Future Research

Although this study has innovatively applied the LDA model to cluster the theme of "World Significance" in the nomination forms of MoW Register, there are still some limitations. First, the existing research couldn't completely eliminate the influence of different length of nomination forms and different writing styles on the research results. How to better explore the commonality of the standard exposition of "world significance" from these differences is worthy of further exploration and research. Second, the research scope needs to be further expanded. This study only analysed "World Significance" criterion of the MoW Programme, leaving out the "Authenticity", "Integrity", and other criteria. Future research should extend a broader horizon to other criteria, or even to the entire review cycle, in order to raise the world's awareness of the need to safeguard documentary heritage. Third, the quantitative research methodology has not been optimised through comparative validation. In this study, only the LDA model was attempted as a technical tool without trying other techniques. Future quantitative research should seek optimum technical path for documentary heritage interpretation by experimenting with a variety of tools, or refining algorithms targeted to the characteristics of documentary heritage.

Acknowledgements

This work is part of Project Number 22ATQ009: "Archival Memory Theory in the Digital Age" funded by The National Social Science Fund of China.

Conflict of interest

No potential conflict of interest relevant to this article was reported.

References

- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Bu, J., & Wang, Y. (2021). *Inherit Human Memory - research on UNESCO Memory of the World Programme*. Soochow University Press.
- Edmondson, R. (2020). Back to the Future: A Reflection on Fundamentals. In R. Edmondson, L. Jordan, & A. C. Prodan (Eds.), *The UNESCO Memory of the World Programme: Key Aspects and Recent Developments* (pp. 311–316). Springer.
- Foster, S., Roslyn, R., Lyall, J., & Marshall, D. (1995). *Memory of the World: general guidelines to safeguard documentary heritage*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000105132?posInSet=5&queryId=f2273d51-7c40-4ff9-abf9-4cc67e70e165>

³ This is an excerpt from the homepage slogan of the official UNESCO website.

- Harvey, R. (2003). UNESCO’s Memory of the World Programme and Australia’s lost and missing documentary heritage. *The Australian Library Journal*, 52(2), 135–147. <https://doi.org/10.1080/00049670.2003.10721523>
- Harvey, R. (2007). UNESCO’s Memory of the World Programme. *Library Trends*, 56(1), 259–274.
- Jordan, L. (2020). Terminology and Criteria of the UNESCO Memory of the World Programme: New Findings and Proposals for Research. In R. Edmondson, L. Jordan, & A. C. Prodan (Eds.), *The UNESCO Memory of the World Programme: Key Aspects and Recent Developments* (pp. 293–308). Springer.
- Kim, K., Park, O., Barr, J., & Yun, H. (2019). Tourists’ shifting perceptions of UNESCO heritage sites: lessons from Jeju Island-South Korea. *Tourism Review*, 74(1), 20–29. <https://doi.org/10.1108/TR-09-2017-0140>
- Kim, S., Lee, H., Jeong, S., & Chung Y. (2021a). Biological distribution and environmental monitoring for the conservation of Janggyeong panjeon Depositories and Daeanggyeongpan (Printing Woodblocks of the Tripitaka Koreana) of Haeinsa Temple in Korea. *International Biodeterioration & Biodegradation*, 156, 105131. <https://doi.org/10.1016/j.ibiod.2020.105131>
- Kim, Y., Son, G., Lee, D., & Son, Y. (2021b). Rural Tourism Image and Major Activity Space in Gochang County Shown in Social Data - focusing on the keyword ‘Gochang-gun Travel. *Journal of Korean Society of Rural Planning*, 27(3), 103–116.
- Suh, K. (2020). History Wars in the Memory of the World: the documents of the Nanjing Massacre and the ‘Comfort Women’. In R. Edmondson, L. Jordan, & A. C. Prodan (Eds.), *The UNESCO Memory of the World Programme: Key Aspects and Recent Developments* (pp. 91–109). Springer.
- Dippon, P., & Helferich, A. (2023). Sharing the documentary heritage of humanity: disparities in distribution. *International Journal of Heritage Studies*, 29(3), 1–22. <https://doi.org/10.1080/13527258.2023.2176347>
- Sievert, C., & Shirley, K. (2014). LDAvis: a method for visualizing and interpreting topics. In J. Chuang, S. Green, M. Hearst, J. Heer, & P. Koehn (Eds.), *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces* (pp. 63–70). Association for Computational Linguistics.
- Springer, J. (2020). Memory of the World: Key Principles and Philosophy. In R. Edmondson, L. Jordan, & A. C. Prodan (Eds.), *The UNESCO Memory of the World Programme: Key Aspects and Recent Developments* (pp. 31–40). Springer.
- Roger, T. (2019). The wonderful UNESCO collection of Panji tales in Leiden University Libraries. *Wacana Journal of the Humanities of Indonesia*, 20(1), 32–55. <https://doi.org/10.17510/wacana.v20i1.746>
- UNESCO. (2011). *Warsaw Declaration*. UNESCO. Drafted on the occasion of the Fourth International UNESCO Memory of the World Conference, 18–21 May 2011, Warsaw. https://archiwa.gov.pl/en/discover/projects/unesco/memory-of-the-world/warsaw-declaration/?doing_wp_cron=1717372807.3291690349578857421875
- UNESCO. (2012). *Memory of the World Register Companion*. UNESCO. <https://www.saac.gov.cn/mowcn/cn/c100450/2021-02/18/4077d201410f4efbb0038431bb29076f/files/f57833f01a2a44d88dae714862375612.pdf>
- UNESCO. (2021a). *General Guidelines of the Memory of the World (MoW) Programme*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000378405>
- UNESCO. (2021b). *Memory of the World Register Companion*. UNESCO. https://www.mext.go.jp/content/20210811-mxt_koktou01-100014749_3.pdf

World Heritage Convention. (2021). *Extended 44th session of the World Heritage Committee: outcomes of the expert meeting on sites associated with recent conflicts and other negative and divisive memories*. UNESCO. <https://whc.unesco.org/archive/2021/whc21-44com-inf8.1-en.pdf>

Notes to table 3

1. “Piri” is from the *Piri Reis World Map*, the earliest record of Columbus’ ocean voyages and the oldest record of the discovery of the New World. It was obtained from the Piri Reis World Map (1513), which was submitted by Turkey and recommended for inclusion in the MoW Register in 2017.
2. “Saiva” means a believer of the Shiva religion, which is an Indian religion and one of the oldest religious beliefs in the world. It was obtained from the *Saiva Manuscript in Pondicherry*, which was submitted by India and recommended for inclusion in the MoW Register in 2005.
3. “Siddhanta” means Siddhanta Saivism, which is one of the main branches of Shivaism. It was also obtained from the *Saiva Manuscript in Pondicherry*.
4. “Utrecht” is the fourth largest city in the Netherlands. It was obtained from the *Utrecht Psalter*, which was submitted by the Netherlands and recommended for inclusion in the MoW Register in 2015.
5. “Fuzuli” is Mahammed Fuzuli (1495–1556), a famous Central Asian Turkic Muslim poet. It was obtained from *The Copy of the Manuscript of Mahammad Fuzuli’s “divan”*, which was submitted by Azerbaijan and recommended for inclusion in the MoW Register in 2017.
6. “David” is a representative Christian figure, the king of Israel, who is mentioned several times in the Utrecht Psalter.
7. “Sa’di” is Sa’di, Moshlefoddin Mosaleh (1208–1291), a Persian poet, who is recognised as one of the four pillars supporting the Persian literary edifice. It was obtained from the *Kulliyāt-i Sa’di*, which was submitted by Iran and recommended for inclusion in the MoW Register in 2015.
8. “Gulistān” is one of the representative works of the poet Sa’di, centring on the core idea of “benevolence”.
9. “Wittgenstein” is Ludwig Wittgenstein (1889–1951), one of the founders of analytical philosophy and one of the most influential philosophers of the 20th century. It was obtained from the *Philosophical Nachlass of Ludwig Wittgenstein*, which was jointly submitted by Austria, Canada, the Netherlands, and the United Kingdom and recommended for inclusion in the MoW Register in 2017.
10. “Borobudur” is the Borobudur Temple Compounds, a ninth-century Mahayana Buddhist temple in Indonesia, which is the world’s largest Buddhist temple and the world’s largest Buddhist monument. It was obtained from the *Borobudur Conservation Archives*, which was submitted by Indonesia and recommended for inclusion in the MoW Register in 2017.
11. “Masaalik” is from the *Al-Masaalik Wa Al-Mamaalik*, which was submitted by Iran and Germany and recommended for inclusion in the MoW Register in 2015. It is one of the most important geographical works of the 4th–10th century AD, and its author Istakhri was one of the founders of geography in the Muslim world.
12. “Bach” is Johann Sebastian Bach (1685–1750), a German composer and keyboard player in The Baroque period, who is known as the “father of Western music”. It was obtained from the *Autograph of h-Moll-Messe (Mass in B minor) by Johann Sebastian Bach*, which was submitted by Germany and recommended for inclusion in the MoW Register in 2015.

13. “Galigo” is from the *La Galigo*, which was submitted by Indonesia and the Netherlands and recommended for inclusion in the MoW Register in 2011.
14. “Nezami” is Ilyas Jamalddin Nezami (1141–1209), a representative poet of Iraqi style in the history of Persian literature. It was obtained from the *Collection of Nezami’s Panj Ganj*, which was submitted by Iran and recommended for inclusion in the MoW Register in 2011.
15. “Khamseh” is the representative work of the poet Nezami.
16. “Dhakhīra” is from the *Dhakhīra-yi Khārazmshāhī*, which is the first medical paper written in Farsi. It was submitted by Iran and recommended for inclusion in the MoW Register in 2013.
17. “Amparo” is from the *Judicial Files Concerning the Birth of a Right: the Effective Remedy as a Contribution of the Mexican Writ of Amparo to the Universal Declaration of Human Rights (UDHR) of 1948*, which was submitted by Mexico and recommended for inclusion in the MoW Register in 2015.
18. “Jacob” is Aletta H. Jacobs, a Dutch doctor and defender of women’s rights. It was obtained from the *Aletta H. Jacobs Papers*, which form the foundation of the international women’s Movement archive. It was submitted by Netherlands and the United States and recommended for inclusion in the MoW Register in 2017.
19. “Constantine” is Learie Constantine, the famous black cricketer who made a prominent contribution towards fighting racial discrimination and easing race relations. It was obtained from the *Constantine Collection*, which was submitted by Trinidad and Tobago and recommended for inclusion in the MoW Register in 2011.
20. “Tripitakan” is from the *Printing woodblocks of the Tripitaka Koreana and Miscellaneous Buddhist scriptures*, which was submitted by Korea and recommended for inclusion in the MoW Register in 2007.
21. “Mcluhan” is Marshall McLuhan, the original 20th century media theorist and thinker. It was obtained from the *Marshall McLuhan: The Archives of the Future*, which was submitted by Canada and recommended for inclusion in the MoW Register in 2017.
22. “Papiamentu” is the language of Papiamentu, which is one of the official languages of the Netherlands Antilles, Aruba, Bonaire, and Curacao. It was obtained from the *First Catechism Written in Papiamentu Language*, which was submitted by the Netherlands Antilles and recommended for inclusion in the MoW Register in 2009.
23. “Diponegoro” was a prince of Yogyakarta-Sudan, Java, Indonesia, and a major leader of the popular uprising against Dutch colonialism in Java, Indonesia, from 1825 to 1830. It was obtained from the *Babad Diponegoro or Autobiographical Chronicle of Prince Diponegoro (1785–1855)*. A Javanese nobleman, Indonesian national hero and pan-Islamist, which was submitted by Indonesia and the Netherlands and recommended for inclusion in the MoW Register in 2013.
24. “Leuven” is the Catholic University of Leuven, an important stronghold of humanism. It was obtained from the *Archives of the University of Leuven (1425–1797): University Heritage of Global Significance*, which was submitted by Belgium and recommended for inclusion in the MoW Register in 2013.
25. “Farquharson” is Charles Farquharson, whose diary records the development of cotton plantations in slavery and non-sugar colonies. It was obtained from the *Farquharson’s Journal*, which was submitted by Bahamas and recommended for inclusion in the MoW Register in 2009.

26. “Bogomil” is from the *Boril’s Synodicon or Synodicon of King Boril*, which takes a negative and critical view of Christian theological doctrine and religious liturgy. It was submitted by Bulgaria and recommended for inclusion in the MoW Register in 2017.
27. “Ilseongnok” is from the *Ilseongnok: Records of Daily Reflections*, which mainly reflects the specific aspects of political and cultural exchanges between the East and West as well as the general patterns of world history from the 18th to the 20th centuries. It was submitted by Korea and recommended for inclusion in the MoW Register in 2011.
28. “Uigwe” is from the *Uigwe: The Royal Protocols of the Joseon Dynasty*, which records the main rituals of the royal family during the Joseon Dynasty from the 17th to the early 20th centuries. It was submitted by Korea and recommended for inclusion in the MoW Register in 2007.
29. “Compendium” is mainly from the *Ben Cao Gang Mu (Compendium of Materia Medica)*, which represents the achievements and development of pharmacology in East Asia before the 16th century and is a milestone in the development of world medicine and culture. It was submitted by China and recommended for inclusion in the MoW Register in 2011.
30. “IICI” stands for International Institute of Intellectual Cooperation. It was obtained from the *Archives of the International Institute of Intellectual Cooperation, 1925–1946*, which was submitted by UNESCO and recommended for inclusion in the MoW Register in 2017.
31. “Bogam” is from the *Donguibogam: Principles and Practice of Eastern Medicine*, which pioneered the world’s concept of state health care for the general public. It was submitted by Korea and recommended for inclusion in the MoW Register in 2009.
32. “Tsunami” is from the *Indian Ocean Tsunami Archives*, which was submitted by Indonesia and Sri Lanka and recommended for inclusion in the MoW Register in 2017.