



Open Library of Humanities

## Representation of Non-Western Cultural Knowledge on Wikipedia: The Case of the Visual Arts

Waqās Ahmed, The Open University, UK, [Waqas.ahmed@open.ac.uk](mailto:Waqas.ahmed@open.ac.uk)

Martin Lewis Poulter, University of Bristol, UK, [m.l.poulter@bris.ac.uk](mailto:m.l.poulter@bris.ac.uk)

---

We explore gaps in Wikipedia’s coverage of the visual arts by comparing the representation of 100 artists and 100 artworks from the Western canon against corresponding sets of notable artists and artworks from non-Western cultures. We measure the coverage of these two sets of topics across Wikipedia as a whole and for its individual language versions. We also compare the coverage for Wikimedia Commons and Wikidata, sister projects of Wikipedia that host digital media and structured data. We show that all these platforms strongly favour the Western canon, giving many times more coverage to Western art. We highlight specific examples of differing coverage of visual art inside and outside the Western canon. We find that European language versions of Wikipedia are generally more “Western” in their coverage and Asian languages more “global,” with interesting exceptions, including that English is one of the most “global.” We suggest how both Wikipedia and the wider cultural sector can address this gap in content and thus give Wikipedia a truly global perspective on the visual arts.

---

Nous explorons des lacunes dans la couverture de Wikipédia des arts visuels, en comparant la représentation de 100 artistes et de 100 œuvres d’art venant du canon occidental avec des artistes et des œuvres d’art notables venant de cultures non-occidentales. Nous mesurons la couverture de ces deux sujets à travers Wikipédia, dans son ensemble et pour chaque version individuelle dans une langue différente. Nous comparons également la couverture concernant Wikimedia Commons et Wikidata, des projets sœurs de Wikipédia qui hébergent des médias numériques et des données structurées. Nous montrons que toutes ces plateformes favorisent le canon occidental, offrant à l’art occidental beaucoup plus de couverture. Nous soulignons des exemples spécifiques de la couverture différente de l’art visuel dans et hors du canon occidental. Nous trouvons que les versions dans des langues européennes sur Wikipédia sont généralement plus « occidentales » dans leur couverture, tandis que celles dans des langues asiatiques sont plus « globales », sauf que des exceptions intéressantes, y compris celle d’anglais, qui est une des versions la plus « globale ». Nous suggérons la façon dont Wikipédia et le secteur culturel élargi peuvent aborder cette lacune de contenu pour donner à Wikipédia une perspective véritablement globale concernant les arts visuels.

---



## **Introduction**

### ***The impact of Wikipedia***

Wikipedia is the world's leading website through which people learn about history and culture. It is the number one informational site on the web and gets many times more use than museum websites. For example, the Metropolitan Museum of Art's (the Met) images on Wikipedia get roughly 10 million hits per month versus 2 million per month on the Met's online catalogue (Maher and Tallon 2018). Each day, there are 260 million views on English Wikipedia from about 70 million users. While it is difficult to know what proportion are for "cultural" articles, it is fair to say there is on English Wikipedia the equivalent of at least one Exposition Universelle (nine million attendance) every single day. The English Wikipedia is just one of nearly three hundred language versions maintained by volunteer communities of differing sizes. The magnitude of this influence brings with it a responsibility of equal measure: to ensure its content is representational of the great diversity of communities and cultures that it engages and informs.

Wikipedia is part of the Wikimedia movement, which includes online platforms, volunteer communities, and charitable organizations, sharing the goals of open knowledge for all. In its current strategy (Meta Contributors 2021a), the Wikimedia movement has explicitly committed to the goal of knowledge equity as one of two core principles: "As a social movement, we will focus our efforts on the knowledge and communities that have been left out by structures of power and privilege." This strategy shapes the grant-making activities of the organizations, and the partnerships sought. For example, Wikimedia's GLAM-Wiki Initiative works with cultural institutions to share their resources openly ("GLAM" is an umbrella term for the cultural heritage sector, encompassing Galleries, Libraries, Archives, and Museums) (Outreach Wiki Contributors 2021b). This includes Wikimedian-in-Residence programmes, in which experienced Wikipedian editors are commissioned by a cultural institution to support an open access culture in the host institution (Meta Contributors 2021c). Although this work is already being done, knowledge equity is such a big task that much more can potentially be done. In this paper we explore how Wikipedia could advance towards knowledge equity in the domain of the visual arts.

### ***Cultural bias***

Various forms of bias on Wikipedia have already been described by research, and a focus of the activity for the Wikimedia organizations is activity to address these biases. Wikipedia's geographic bias and gender bias have their own literatures, so will be outside the scope of the present research. Here we focus specifically on cultural bias,

that is, underrepresentation or misrepresentation of aspects of the cultures of the non-Western world. It has long been observed in the literature that the different language versions of Wikipedia reflect cultural biases of, and celebrate the “local heroes” of, their respective language communities (Callahan and Herring 2011; Maurer and Kolbitsch 2006). For example, the biographies in European-language Wikipedias do not follow the pattern of world population but greatly emphasize the culture of Western Europe and the United States (Graham, Hale, and Stephens 2011).

Cultural biases existing on Wikipedia can generally be considered a reflection (both a cause and a consequence) of biases existing in the literature and more widely in society. These societal biases have a long and well-documented history, rooted in systems of hegemony and oppression like imperialism. Seminal works such as Edward Said’s *Culture and Imperialism* have spotlighted how many of these biases persist in the postcolonial era. Globalization facilitated less of a proportionate cross-cultural exchange and more a spread of the predominant culture (that is, Western).

The term “art” has a complex history (Steiner 1996; Dean 2006), and we recognize that defining it in general is a tricky business, especially in the context of recognizing different cultural qualifications. Universalizing the term, which to some extent must be done for the purpose of comparative analysis, comes with the risk of potentially “employing a Western bias to explore a Western bias,” thus replicating the bias. Our attempt to minimize such a risk is outlined in the next section of this paper, where our respect for definitions and hierarchies is reflected in the inclusion of various works that are considered “visual art” according to various non-Western cultural traditions.

The internet initially promised to make geography irrelevant, but algorithms have created new kinds of inequality in the amount of data about physical locations and its availability to different language communities (Graham and Zook 2013). Recent activism, such as Black Lives Matter and the debate over the holdings of European museums, has underlined the urgency of unearthing overlooked or oppressed histories and cultures. These questions are being raised in the most traditional cultural institutions, as well as by online platforms such as Wikipedia.

### ***The visual arts***

Whereas many forms of bias relating to a specific culture—such as its music, language, literature, performing arts, history, fashion, food, philosophical ideas—clearly exist, this paper pertains specifically to the visual arts. As per the scope of this paper, the culture under examination is the entire “non-Western” world (a concept defined later).

A pro-Western cultural bias relating to the visual arts can be demonstrated with a superficial survey of visual-art-related lists on English Wikipedia, the largest language

version. For example, its “list of sculptors” is 99% Western, its “list of painters by nationality” is around 75% European, and its “list of contemporary visual artists” is 80% European. Moreover, many countries (even those with especially rich artistic traditions, such as Libya and Mali) do not even have dedicated articles about their art in the same way that there exist exhaustive articles such as “Art of France” or “Art of Greece.” This national bias is further evidenced by the “list of national museums” where non-Western national museums (even those among the most visited in the world, e.g., Brazil) have relatively short, insufficient articles, often without collections galleries (something that is almost a given for most major Western museums). It is also indicated by the fact that despite there being many museums in the non-West dedicated to a single artist, the articles covering the “list of single artist museums” and “museums devoted to one artist” are 90% Western.

One could imagine a situation where Persian Wikipedia had a similar emphasis on Middle Eastern art and so on: in other words, where these imbalances in coverage were all due to the “local hero” effect. Instead, we think a larger bias is at play. Our hypothesis is that Wikipedia (taking all its language versions as a whole) has significant and systemic imbalances in the representation of non-Western visual arts, and that these can be identified and addressed. As such, the main objectives of this research are: to identify those areas in Wikipedia’s coverage of the (visual) arts where there are significant imbalances according to culture, language, and geography; to ascertain the scale and nature of these imbalances; to describe what a more equitable representation of visual arts on Wikipedia would look like; and finally, to suggest strategic and practical ways towards that greater balance, building on the work already being done by the Wikipedia communities and organizations.

### ***Paper structure***

To test the hypothesis concerning the representation of non-Western cultural content on Wikipedia, this paper will take both a quantitative and qualitative approach. A research methodology based on making comparisons of the coverage of Western artists and artworks vis-à-vis their non-Western counterparts will be employed.

- Identifying 100 leading Western artists, assessing the extent and quality of their coverage in English and other languages
- Identifying 100 leading non-Western artists of comparable calibre/stature—assessing the extent and quality of their coverage in English and other languages
- Making a comparison and drawing out several case studies as examples

- Identifying 100 leading Western masterpieces—assessing the extent and quality of their coverage in English and other languages
- Identifying 100 leading non-Western masterpieces of comparable calibre/stature—assessing the extent and quality of their coverage in English and other languages
- Assessing the variation of imbalance according to the platform (Commons versus Wikipedia versions versus Wikidata)

## Methods

### *Definitions and scope*

What exactly are we classifying as “visual art”? In theory, visual art can refer to a range of artistic expressions including conceptual art, installation art, and contemporary art, but this paper will focus on the traditional art forms that have been practiced over the centuries and across the world and have often been referred to as “fine art.” Yet what is considered “fine art,” too, differs according to different cultures: The hierarchy in the West has placed epic easel painting at the highest, whereas in the Islamic world calligraphy is among the highest, as are textiles and miniatures in Persia and calligraphic landscapes in China, and in Japan there is a special reverence for decorative and applied arts.

This study balances the need to be respectful to each of these hierarchies whilst also standardizing to some degree to allow for reasonable comparison. After careful consideration of these cultural sensitivities, it was decided that the paper should largely focus on painting and sculpture but also include other media such as illuminated manuscripts, textiles, and calligraphy. It does not include architectural features, although it must be noted that much artistry and craftsmanship—for example, the stained-glass windows of European Cathedrals or the geometric tilework and calligraphic inscriptions in Samarkand, Bukhara, and Alhambra—was recruited to serve aims of aesthetic creativity. The study does not include architecture, ancient artifacts, manuscripts (unless with calligraphy and illumination of considerable merit), jewellery, furniture, or fashion.

Many of the artists involved in these projects—particularly outside the West—remain anonymous.

The “West” is a problematic term and concept, as it promotes the notion of a bipolar, dichotomic world. What is classified as non-Western culture is all culture originating and prevailing outside of Europe, Scandinavia, Russia, Eastern Europe, North America, and Australasia, except for those cultures (now in the minority) indigenous to those lands, such as aboriginal and Inuit. This is an extremely large group.

Is it fair to put Europe with a population of one sixth of the world against the rest? It would in theory be more apt to compare Europe with another continent such as Latin America or Africa. This should be an absurd exercise, but in fact the results show it is absurd for exactly the opposite reasons.

The time scope of art in this study is roughly 1000 years. There are many reasons for this. Firstly, this covers the emergence of the conventional East-West dichotomy, and therefore the “West and the rest” narrative that continues to this day. Secondly, this period comprises major cultural civilizations from across the world and therefore various artistic golden ages, which celebrated, commissioned, recorded, and preserved the works of leading artists. Thirdly, this covers the era of the great European empires, which collectively governed the majority of the non-European world—important, as (especially) the last 500 years of European colonialism suppressed or looted many indigenous works from the colonies, the legacy of which is much of the knowledge imbalance that this paper seeks to highlight. Fourthly, before this period, artworks were often considered artifacts (or sometimes in the Western case antiquities) rather than masterpieces produced by an individual artist, or even a guild or atelier. A typical demonstration of this might be the exhibition of the piece in a historical museum rather than a dedicated fine art gallery.

#### ***Identifying Western artists***

English Wikipedia has a system of “Vital Article” lists that define topics judged to have different levels of encyclopedic importance (Wikipedia Contributors 2020). Level 1 contains ten articles (including “The Arts”), Level 2 contains one hundred articles (including “Visual Arts”), and so on. These lists are compiled irrespective of the quality of the existing articles. It is fair to the Wikipedia community to use a standard they have set themselves, so we took the Vital Article lists as a starting point.

The 10,000 topics at Vital Article Level 4, as of November 2020, included 78 Western artists; our shortlist began with these. The additional 22 artists were selected after consultation with the wide range of lists available in media articles and published books. “Top 100 Artists” lists are common with regards to Western artists. In our choices we aimed to diversify a list dominated by painters from a few European countries, introducing women, decorative artists, and Scandinavian artists.

#### ***Identifying non-Western artists***

The same methodology for establishing the set of leading non-Western artists was simply impossible. For instance, only three leading non-Western artists have vital



articles (Hokusai, Riviera, and Kahlo). No single definitive list exists as a counterpart to the abundance of sources defining the Western canon. Therefore, a mixed methodology was developed towards making a list of 100 artists that could credibly serve as a counterpart to our Western list.

One of the starting points was to consult the lists already available on Wikipedia. The “list of African artists” and “list of Chinese artists,” for example, provided a sound basis for further investigation, as it is these lists—however inadequate—that we intend to amend and enrich as a result of the research. This initial compilation of non-Western artists was then cross-referenced against those listed through Google search’s respective lists such as “African artists” or “Chinese artists.” As Wikipedia and Google lists of this sort are usually considered indicators of popularity, those appearing on both lists were shortlisted for further investigation.

Separately, a digital media search was conducted, and a number of magazine articles, for example, “Top African Artists” or “The Greatest Japanese Sculptors” and other such rankings were consulted. Where names appeared frequently in different articles, those were shortlisted and again cross-referenced with existing lists. A high-level (though limited) literature review of books and articles was conducted to list the canon in each major region according to academic experts. These were again cross-referenced against existing lists with a view to shortlisting those artists who were both popular as well as critically acclaimed.

Another measure or “marker” for artists deserving a place on this shortlist was whether they had attained official recognition through national and international awards, as well as receiving the highest national honours for their contribution to visual arts, as well as those considered “national artists” or those appointed “imperial court artists.” Some of these names overlapped with existing research whereas others required further validation. Much of this validation came from interviews with experts in the respective fields of art. These experts are listed in the Acknowledgements.

Finally, we cut down the lists of Western and non-Western artists to make lists that were similar in terms of time-period coverage and were diverse in multiple respects. It is important to note here that the resultant list (in Table 8 and Table 9 [Appendix A and Appendix B]) is a representative and indicative sample, sufficient for this particular study to test the hypothesis and provide indicative results. It is not exhaustive and certainly not aimed at establishing a definitive “Top 100.” The latter would be outside the scope of this paper and require extensive research and consultation, warranting a paper in its own right.

The English Wikipedia defines a topic as notable when it has significant coverage in at least three reliable sources. Language versions of Wikipedia differ somewhat in their notability standards. All the artists identified through the various forms of research can be considered notable, and therefore deserving of Wikipedia articles. For the purposes of this study, where the objective was to have a representative sample list of counterpart artists to those in Western culture, shortlisting through this process of verification suffices. Some names who created more than one masterpiece were also included.

### ***Identifying Western masterpieces***

As with the Western artists, we used English Wikipedia's lists of Vital Articles as a starting point for our target list of masterpieces. Getting the relevant articles from Vital Articles Level 5 and filtering out some that were ancient or too recent gave us 170 works. Wikidata allowed us to identify that 78 of these works had articles in Encyclopedia Britannica, which was an additional cue to notability. The longlist included many cases of multiple works by the same artist, so we cut this list down to 100 while preserving diversity by removing works by artists who were already included (see Table 10 [Appendix C]).

### ***Identifying non-Western masterpieces***

The process of shortlisting a representative set for leading non-Western masterpieces was different from all of the above, though there are some similarities with the process of researching non-Western artists.

This list was the most challenging to compile; firstly, this is because no such list currently exists, and secondly because substantial research into non-Western masterpieces would simply unveil too many options to shortlist from. Though Wikipedia and Google search unearthed some notable examples of non-Western masterpieces, this method was not as helpful as it was for researching non-Western artists.

So, we began by including the most celebrated works listed as "national treasures" by various non-Western countries, namely those that subscribed to our remit of visual art. In addition, highlights from National Museum and Galleries collections across Asia, Africa, and Latin America were also longlisted, as were those identified from a media review as artworks of symbolic significance or representing an important cultural movement. We added to this a select number of works from the non-West that broke sales records at major auction houses, as well as names appearing repeatedly through our literature review. The list was finalized after cross-referencing with scholarly



experts and shortlisted to 100 based on the expert discretion of the authors of this paper (see Table 11 [Appendix D]).

### ***Quantitative comparison***

The finalized lists of Western and non-Western artists and masterpieces defined four content areas whose coverage we could explore both quantitatively and qualitatively.

The Wikimedia family of sites allows users to build, remix, and share open content about visual art in different modalities. We measured how three different platforms represent the topics on our Western and non-Western target lists.

On Wikipedia, there are narrative articles. On Wikimedia Commons, there are freely licensed images and other digital media with associated metadata. The images are used to illustrate Wikipedia articles and other educational materials and constitute an educational and research resource in their own right. On Wikidata, there are machine-readable statements (such as that Auguste Rodin was born in Paris) with attached citations. These statements can be extracted by custom queries and visualizations and are used in applications inside and outside Wikimedia. These include the “infoboxes” that give basic facts about a topic in a Wikipedia article or Commons category index. There are other Wikimedia platforms, but just these three—the most relevant to visual arts—are considered in this paper. Wikipedia exists in hundreds of different language versions, while Wikidata and Commons are each single, multilingual sites.

A Wikipedia article can be anything from a single line of text to a 20,000-word essay. A minimal Wikidata representation of an artist consists of a name, a one-line description, basic statements (e.g., this is [1] a human being, [2] of male gender, [3] whose occupation is sculptor), and perhaps an authority file identifier. A more fully developed Wikidata representation will include dozens of biographical details, including family relations, places of education and work, and identifiers in potentially hundreds of external sites and databases. So, when measuring the representation of the topic, it is important to account for the size of the article or data item, not just its presence or absence.

The Wikimedia sites have APIs (Application Programming Interfaces) that allow external code to request specific information such as the length of an article (MediaWiki Contributors 2021). In the case of Wikidata, these can include sophisticated database queries. We wrote code that, via the APIs of Wikidata, Commons, and the many different language versions of Wikipedia, extracted the quantitative information needed for our target lists.

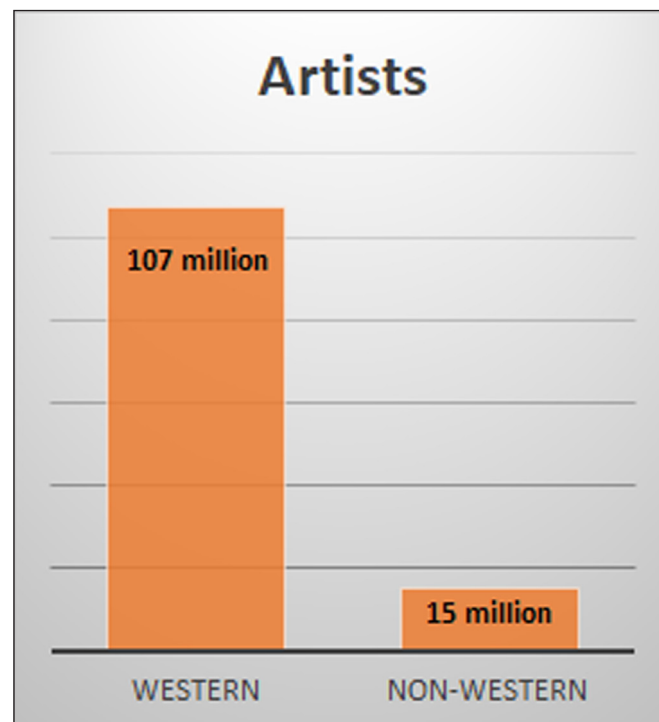
## Results

### *Quantitative analysis*

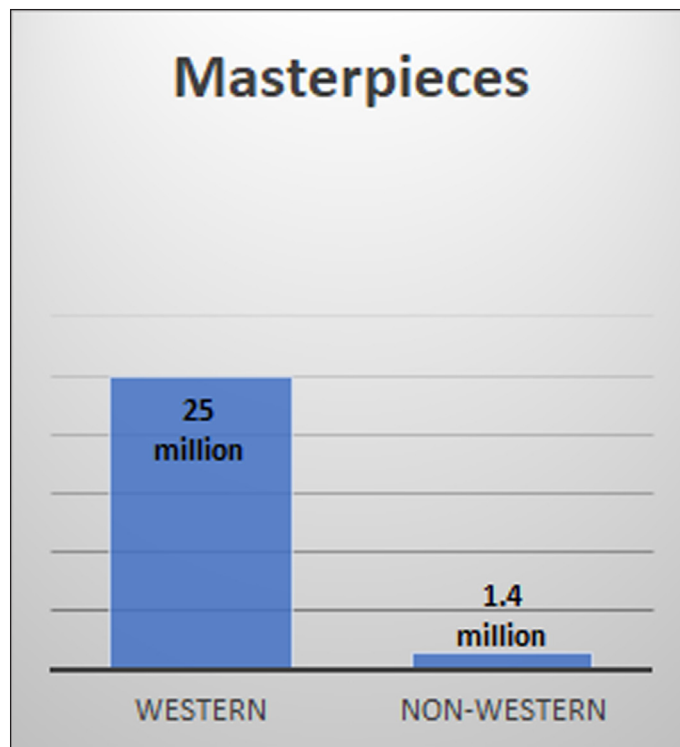
#### Wikipedia articles

Wikidata queries provide all the Wikipedia articles about a given topic—in this case, articles about the artists and artworks in our lists. Our code then requested the byte length of each article from the relevant language version of Wikipedia. Byte length is a fairer measure of the content of an article than character count. For example, characters in English take one byte each, in Hebrew two bytes each, and in Chinese three or four bytes each.

It was discovered that there were five times as many articles about our Western artists (total 7,808) as non-Western (1,621) and sixteen times as many for Western masterpieces (2,570) as for non-Western (165). The most-represented artist, Leonardo da Vinci, has articles in 222 language versions of Wikipedia. Taking article size into account, there is a little over seven times as much Wikipedia coverage of the Western artists (107 million bytes) as non-Western (15 million), and eighteen times as much of the Western masterpieces (25 million) as non-Western (1.4 million) (see **Figures 1 and 2**).



**Figure 1:** Total article size (bytes), across all versions of Wikipedia, for the artists on our lists.

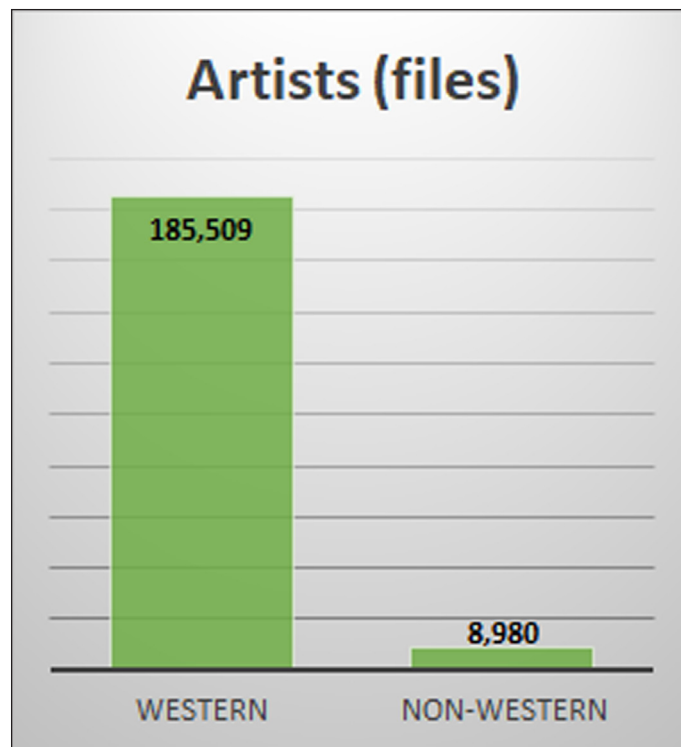


**Figure 2:** Total article size (bytes), across all versions of Wikipedia, for the masterpieces on our lists.

### Digital media files

Files on Wikimedia Commons can be tagged with an artist's name for many reasons. They may be a depiction of that artist, a photograph of an artwork, or a document relating to them. The connection can be more tenuous: photographs of places where the artist lived, or of places named after them. A Wikidata query provided us with the categories relating to our chosen artists. Categories can contain sub-categories, and so on iteratively, so to get total numbers of files we used the Commons API and, for a few especially large categories, the PetScan tool created by Magnus Manske (<https://petscan.wmflabs.org/>). There might be files related to a topic that exist on Commons but are not categorized appropriately, or where the category link exists but is not known to Wikidata, so our measure might underestimate the coverage of obscure topics, although we mitigated this by searching directly on Commons and adding a few links that were missing in Wikidata.

We found twenty-one times as many files for Western artists (total 185,509) as for non-Western (8,980 files). All of the Western artists had a category on Commons compared to 84 of the non-Western (see **Figure 3**).



**Figure 3:** Total numbers of files available on Wikimedia Commons for the artists on our lists.

#### Database statements

On Wikidata, all of our Western artists and masterpieces were already represented. Of the 100 non-Western artists, 99 already existed in Wikidata, along with 34 of the 100 non-Western masterpieces. Wikidata's query service allowed us to count the statements for each. We found just under four times as many statements about Western artists as non-Western artists, and nine times as many statements about Western as non-Western masterpieces (see **Figures 4** and **5**).

#### Differences across language versions

The language versions of Wikipedia have contributor communities that vary greatly in their size and where they are located. Thus, they vary in the amount of text they have produced and about what topics. For each pair of an artist and a language version of Wikipedia, our data have a byte count expressing the size of the artists' article in that language. By summing across each language, we can compare our matched lists, measuring the degree to which different Wikipedias prioritize the Western canon in the field of visual arts. Since we are comparing the coverage given to matched lists, our measure is not directly affected by the size of the Wikipedia itself.

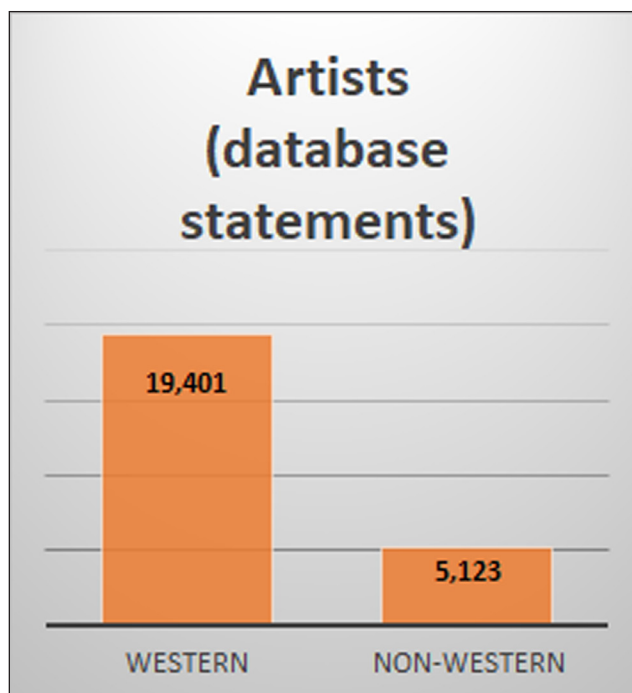


Figure 4: Total number of statements in Wikidata for the artists on our lists.

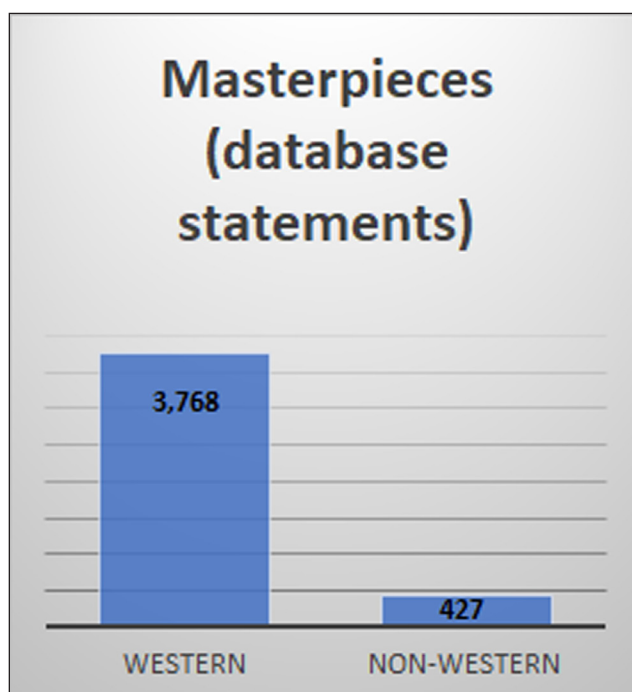


Figure 5: Total number of statements in Wikidata for the masterpieces on our lists.

Our measure is each Wikipedia's coverage of our Western artists divided by its coverage of the non-Western artists. Thus, higher numbers mean a more Western focus and lower mean more global. **Table 1** shows this ratio for 86 of the larger Wikipedias. Six of them give more coverage to our non-Western than to Western artists.

Language	Language code	Western artists (bytes)	Non-Western artists (bytes)	Ratio
Thai	th	1577064	37777	41.75
Galician	gl	1560347	106539	14.65
Italian	it	3846109	279501	13.76
Serbian	sr	1877184	147803	12.70
Polish	pl	1856378	157538	11.78
Simple English	en-simple	478046	43888	10.89
Hungarian	hu	1553127	152783	10.17
Hebrew	he	1243742	137012	9.08
Turkish	tr	1130276	133821	8.45
Portuguese	pt	1822828	216634	8.41
Japanese	ja	2893884	344564	8.40
Czech	cs	1685339	204217	8.25
German	de	4513825	555219	8.13
Spanish	es	4202760	517974	8.11
Dutch	nl	1636039	228762	7.15
French	fr	6235180	876287	7.12
Malayalam	ml	1374045	205898	6.67
Catalan	ca	2254534	341903	6.59
Welsh	cy	391141	68955	5.67
Russian	ru	5330034	958510	5.56
Vietnamese	vi	1042035	194942	5.35
Chinese	zh	1274771	242399	5.26
Arabic	ar	1401071	267791	5.23
Ukrainian	uk	2797317	614581	4.55
Armenian	hy	2239028	530142	4.22
Persian	fa	1634226	392738	4.16

(Contd.)



Language	Language code	Western artists (bytes)	Non-Western artists (bytes)	Ratio
English	en	5927835	1494254	3.97
Indonesian	id	565859	171585	3.30
Hindi	hi	337095	121612	2.77
Punjabi	pa	340529	177997	1.91
Bengali	bn	598682	343743	1.74
Gujarati language	gu	221480	165210	1.34
Urdu	ur	65583	110819	0.59

**Table 1:** Coverage for Western and non-Western artists in some larger language versions of Wikipedia. The full version of this table is given in Table 12 (Appendix E).

As expected, European languages tend to have higher ratios while Asian languages are lower. There are anomalies; Thai is the most Western in its coverage of visual arts, and English and Scots are among the most global. The ratio across all Wikipedias is, as we have seen above, just over 7. So Japanese Wikipedia, with a ratio more than 8, is more focused on the Western canon than the Wikipedias as a whole.

### *Comparative examples*

Having explored the size of the content gap, we now illustrate it with specific examples of artists, artworks, and art movements.

The Sistine Chapel in the Vatican and the Sultan Ahmed Mosque (Blue Mosque) in Istanbul are two of the world's most visited places of worship—each having approximately 5 million visitors a year, making them comparable in terms of places of considerable interest to devotees and to tourists. Importantly, interest in them is not only because they are places of religious and historical significance, but also because the interiors of these places are considered to be of works of tremendous artistic merit. This is particularly the case with their ceilings. The ceiling of the Sistine Chapel was painted by the master Michelangelo in the early sixteenth century and is itself considered an iconic masterpiece in the history of Western art. It is composed of various Biblical stories painted in traditional Renaissance figurative style. The ceiling of the Blue Mosque was likewise painted by a master, though in this case, the master calligrapher Syed Kasim Gubari. Like Michelangelo, Gubari is considered one of the great masters in the history of his region/culture's art (in this case, Ottoman/ Islamic).

Whereas Michelangelo is extensively represented on Wikipedia (3,902,976 bytes in 198 language versions), Gubari has minimal representation (short articles in four languages, totalling 8,772 bytes). Moreover, “Sistine Chapel Ceiling” has an extensive Wikipedia article whereas “Blue Mosque Ceiling” does not have an article or even a Wikidata entry (see **Table 2**).

	<b>Sistine Chapel ceiling</b>	<b>Blue Mosque ceiling</b>
Wikipedia	936,019 bytes in 25 languages	n/a
Wikimedia Commons	597 files	253 files (Category: Interior of Sultan Ahmed I Mosque)
Wikidata	52 statements	n/a

**Table 2:** Coverage of a significant artwork from Christian culture and Islamic culture on the Wikimedia projects.

Su Shi, the 11th-century Chinese artist whose painting broke the record for highest selling Asian artwork, was a polymath also celebrated as a poet, engineer, litterateur, scientist, and political figure. He is covered in 35 language versions of Wikipedia, whereas the Western polymath and comparably versatile artist Leonardo da Vinci is one of the most covered artists on Wikipedia, with articles in 222 languages totalling nearly five million bytes (see **Table 3**).

	<b>Leonardo da Vinci</b>	<b>Su Shi</b>
Wikipedia	4,823,238 bytes in 222 languages	328,858 bytes in 35 languages
Wikimedia Commons	23,164 files	267 files
Wikidata	376 statements	120 statements

**Table 3:** Coverage for a prominent European artist/polymath and a prominent Asian artist/polymath on Wikimedia projects.

Likewise, comparably celebrated royal court portrait painters such as Hans Holbein (15th-century England) and Mihr Ali (18th-century Persia) have remarkably different Wikipedia coverage levels (see **Table 4**).

	<b>Hans Holbein</b>	<b>Mihr 'Ali</b>
Wikipedia	854,397 bytes in 63 languages	40,854 bytes in 6 languages
Wikimedia Commons	2,232 files	21 files
Wikidata	205 statements	21 statements

**Table 4:** Coverage of two royal court painters on the Wikimedia projects.

Beyond artists and artworks, another way of seeing the disproportionality in representation of the visual arts is by analyzing Western artistic movements vis-à-vis counterparts outside the West. For example, the Pre-Raphaelite Brotherhood in 19th-century England was a major movement that sought a return to traditional forms of Western art and comprised a number of notable artists, critics, and patrons (such as Millais, Burne-Jones, Gabriel-Rossetti, Ruskin, Morris). It is extensively covered on Wikipedia, Commons, and Wikidata. The Bengal School of Art likewise rejected modernism and sought a reversion to traditional forms, and also included major artists, critics, and patrons such as Bose, Tagore, and Kastghir. Its coverage on Wikipedia is minimal in comparison to that of the Pre-Raphaelite Brotherhood (see **Table 5**).

	<b>Pre-Raphaelite Brotherhood</b>	<b>Bengal School</b>
Wikipedia	876,061 bytes in 54 languages	31,148 bytes in 3 languages (English, French, Bengali)
Wikimedia Commons	10,233 files	121 files
Wikidata	43 statements	5 statements

**Table 5:** Coverage of a Western artistic movement and a non-Western artistic movement on the Wikimedia projects.

Another suitable comparison might be the European Post-Impressionists and the Japanese Nihonga movement (see **Table 6**).

	<b>Post-Impressionism</b>	<b>Nihonga</b>
Wikipedia	407,327 bytes in 65 languages	136,979 bytes in 16 languages
Wikimedia Commons	31,041 files	2,322 files
Wikidata	42 statements	13 statements

**Table 6:** Coverage of a Western artistic movement and a non-Western artistic movement.

## Discussion

We have replicated the common finding of a “local hero” effect, with European artists given higher priority in European-language Wikipedias, but that is not the most salient result. Looking at Wikipedia as a whole, and at the multilingual sites Wikidata and Wikimedia Commons, we found large differences in their relative coverage of our Western and non-Western artists: ratios of 7, 4, and 21, respectively. We showed earlier that an examination of English Wikipedia shows a strong emphasis on Western rather than non-Western art; it turns out that English is one of the least biased major Wikipedias in this respect.

Wikipedia’s volunteer contributors summarize published sources, including books, research papers, and institutional catalogues. Factors that might contribute to an imbalance of coverage include the extent to which different kinds of art are described in published sources, the availability of those sources to Wikipedia contributors (in their language and in forms that they can access), and the interests and priorities of contributors to a given language version.

By our quantitative measure, Wikidata has much less Western bias than Wikipedia collectively, and Wikimedia Commons has much more. The differences in ratio for different platforms can be understood in terms of how each platform sets floors or ceilings on the size of representations. Wikimedia Commons has no upper limit on the number of digital files that can be tagged with a given topic. While there is no technical upper limit on the statements about a topic in Wikidata, there are only a certain number of properties that can be represented in that database. Wikipedia’s style guides put upper limits on the length of articles—usually that they should not exceed 100,000 bytes—although these can vary between languages and are not rigidly enforced.

That English Wikipedia is relatively balanced compared to other language versions (but still giving a small fraction of coverage to the non-Western artists) might be due to the great deal of scholarship being published in English and research done in English-language institutions. It might reflect the activity of Wikimedia chapters and groups that have built partnerships with cultural organizations. It could conceivably be a ceiling effect from its being the largest Wikipedia. If the Western canon is already as extensively documented as it can be, an English Wikipedia contributor wanting to create a new article about an artist is more likely to look to non-Western topics.

If, instead of ratios, we consider the absolute size of coverage of non-Western art, we see that this coverage is most extensive on European-language Wikipedias. The languages that have more than 500,000 bytes of content about non-Western artists are German, Spanish, French, Russian, Ukrainian, and Armenian. This is unsurprising given that these are among the largest versions of Wikipedia, with relatively large volunteer communities. It suggests one interim way to address the imbalance and make other language Wikipedias more global may be to translate articles from these to other languages. This, ironically, would help improve the pro-European emphasis of Wikipedia as a whole, although it would mean that the articles are drawn primarily from sources in European languages. This would be a step in the right direction, but not a solution to the problem of knowledge inequity due to systems of power and privilege for which we suggest bolder action later on.

We did a follow-up analysis focusing on coverage of the Arabic and Persian artists and masterpieces. Summing the coverage of these topics, excluding those languages whose total coverage is less than 100,000 bytes, gives us the results shown in **Table 7**.

This underlines that, although Russian (the seventh-largest Wikipedia) gives a small proportion of its coverage to our non-Western art when compared to Western (a ratio of 5.6), its sheer size means that it has more content about Arabic and Persian visual arts than Arabic Wikipedia does. Hence, it would help Wikipedia become more global by our blunt quantitative criterion if there were translations of articles from Russian or English to Arabic.

Lang. code	Language	Total article size (bytes)
fa	Persian	332,648
en	English	326,158
cy	Welsh	234,738
ru	Russian	229,462
ar	Arabic	169,081
fr	French	155,712
de	German	128,296
es	Spanish	127,264

**Table 7:** Total coverage in some language versions of Wikipedia for Arabic and Persian artists and masterpieces from our lists.

One way to bulk-create articles is to paste biographical tables (name, dates, fields of work) from Wikidata into a textual template. This is used by the Reasonator tool to generate very short biographical profiles. One Reasonator entry reads: “Kawade Shibatarō was a Japanese artist. He was born in 1856. His field of work included cloisonné. He died in 1921” —with some links to a few examples of his work (Wikidata Contributors 2022). Welsh Wikipedia has deployed a similar process, which accounts for its extensive coverage despite having a relatively small community of volunteer contributors. While articles created this way lack the narrative nuance of a human-written article, they give basic facts about a topic and have automatically generated citations. This demonstrates another way Wikipedias can build their coverage of an under-represented topic.

## Conclusions and recommendations

### *Recommendations for the cultural sector*

The representation of a topic on the Wikimedia sites depends on multiple factors. Suitable sources need to be available; suitably licensed images need to be uploaded or put where Wikimedia volunteers can easily access them; and the writing, reviewing, and improvement of a Wikipedia article takes effort. Organizations such as museums,

galleries, and publishers can thus help extend the representation of non-Western art in various ways.

- Paywalled publications are a significant barrier for most Wikipedia contributors, so it is helpful if existing research can be put on open access.
- One way to kickstart Wikipedia articles is by repurposing existing text publications. These need to match Wikipedia's purpose by summarizing mainstream scholarship on a topic rather than reporting new research or synthesis, and they need to be freely licensed. Such articles can be pasted into Wikipedia and given an attribution template that credits and links the original source (Wikipedia Contributors 2021a).
- Wikipedia is a summary of reliable sources, and increasing the range of sources about non-Western art would serve the knowledge equity goal of "sharing knowledge [...] left out by structures of power and privilege." The implicit knowledge of experts was crucial to our research, and more of this implicit knowledge could be made explicit by being published.
- Image collections, whether out of copyright or freely licensed, can be shared by direct upload to Wikimedia Commons or at least placed openly online where Wikimedia volunteers can access them. There are tools and processes for doing this in bulk and for making sure the files have suitable metadata (Wikimedia Commons Contributors 2021).
- By employing a Wikimedian in Residence, an institution makes the best use of Wikimedia platforms to ensure the visibility of its collections. An experienced Wikimedia contributor will be able to make images findable, engage a wider community, and report on metrics of success. Wikimedia's local chapters can help institutions recruit suitable Wikimedians (MediaWiki Contributors 2021).
- Cultural institutions can also provide identifiers and basic biographical information for artists and works, which can be linked from Wikidata and used to establish notability.
- The Wikipedia Library (Orlowitz 2018) is an initiative in which publishers of paywalled scholarship can give temporary access to selected Wikipedia contributors, helping them create and improve articles with citations to those scholarly sources. Publishers of relevant material can consider joining this if open access is not an option. Oxford Art Online, published by the Oxford University Press, is a relevant source available through this method, which more publishers could adopt.



- The OpenGLAM Principles (OpenGLAM Working Group 2011) set out how a cultural institution can use its intellectual property policy and technical infrastructure to promote the widest engagement with its collections. The principles, currently being revised, capture actions that would be helpful to the Wikimedia platforms as well as the wider community.

### ***Recommendations for the Wikimedia contributor communities***

Wikipedia and Wikimedia volunteer contributors can take action straight away to reduce the content gaps described in this paper.

- An outstanding example of work to reduce a content gap on Wikipedia is the Women in Red project (Wikipedia Contributors 2021b). This addresses the gender content gap by using Wikidata and other sources to build “redlists”: lists of notable women who do not yet have a Wikipedia article and whose links are therefore red. In addition to these lists of target articles, the Women in Red project pages include bibliographic sources, guidance, and supporting materials for “editathon” events. Volunteers can choose an article to create, turning the link from red to blue. During the existence of the project, tens of thousands of new English articles relevant to women have been created. It is hard to know how much new content to attribute to a specific effort, but research has found a rise in article quality for the broad topic of women scientists compared to articles in general (Halfaker 2017). We propose that there should be similar projects for the gaps in representation of the visual arts. The Wikidata identifiers and other information in our appendices can be used to make redlists.
- The community should consider adding artists and masterpieces from our non-Western lists to the Vital Article lists on English Wikipedia and any counterparts on other language versions.
- Since 2015, Wikipedia has had a Content Translation tool, which prepares a machine-translated version of an article that a human user can correct and publish (Dolmaya 2017). We have seen that English, French, and Russian Wikipedias have a relatively large volume of coverage of non-Western art, so translation of those articles into more languages would improve the balance.
- A crucial supply of Commons images comes from photographs of out-of-copyright works that museum visitors have taken and then uploaded. For museums that do not have a formal programme of digitization, this informal digitization is an option for creating digital content. It requires the institution

to allow, even encourage, visitors to take photographs as part of their engagement during the visit.

### ***Recommendations for Wikimedia organizations***

Addressing knowledge gaps is already a main focus of the activity of the Wikimedia organizations (the San Francisco-based Wikimedia Foundation and the national and thematic Wikimedia Chapters). This takes the form of supporting or enabling community activities described in the previous section; funding dedicated research, software, or outreach; or building partnerships with other organizations (Meta Contributors 2021b).

The list of existing cultural partnerships shows that Wikimedia has been successful in Europe and North American in building cultural partnerships with major institutions such as the Metropolitan Museum of Art and the British Library. There are many national institutions in the rest of the world that have not had any kind of partnership (Outreach Wiki Contributors 2021a). When looking on Commons for partnerships that had shared Islamic calligraphy, we found the Met, the Cleveland Museum of Art, the Library of Congress, Los Angeles County Museum of Art, and the National Library of Israel. So, the material Wikimedians are working with to document Islamic art is coming mostly from the United States and not from institutions in the Islamic world. To address the gap described in this paper, the Wikimedia organizations should seek partnerships with national as well as grassroots cultural institutions across Asia, Africa, and Latin America, as well as indigenous communities across North America and Oceania. We provide a list of relevant institutions in Table 13 (Appendix F).

### ***Limitations and further research***

Further subdivisions of the categories of Western and non-Western art and artists offer additional research questions that could be investigated. For example, examining gender parity in the history of Western art vis-à-vis the history of non-Western art in Wikipedia was outside the scope of this study, but clearly emerged as an important and necessary area of further research. Also related specifically to representation on Wikipedia, investigating the extent to which disproportionality in such content related to racial, ethnic, geographical, cultural, and religious disproportionality in editors and readers would also be important.

Perhaps more indirectly related to representation on Wikimedia, investigating people's general knowledge of non-Western art history and exposing the bias or

ignorance even among those considered to be “cultured” or reasonably knowledgeable about art history (such as students and scholars) would be helpful in explaining how this is reflected on Wikipedia.

---

### **Additional File**

The additional file for this article can be found as follows:

- **Appendices.** Appendix A (Table 8) to Appendix F (Table 13). DOI: <https://doi.org/10.16995/dscn.8078.s1>

### **Acknowledgements**

This research was supported by a grant from Wikimedia UK. We are grateful to Daria Cybulska (Wikimedia UK's Director of Programmes and Evaluation) and two anonymous reviewers for comments on previous drafts of this paper. Any errors or omissions are the fault of the authors.

### **Experts Consulted**

Professor Christian Luczanits, SOAS, University of London, UK – Himalayan Art

Professor McCausland, SOAS, University of London, UK – Chinese and East Asian Art

Professor Chika Okeke-Agulu, Princeton University, USA – African art

Professor Crispin Branfoot, SOAS, University of London, UK – Indian art

Professor Sir Nasser David Khalili, The Khalili Collections, UK – Islamic and Japanese art

Professor Maria Madero, London Interdisciplinary School, UK – Latin American art

Dr. Heather Igloliorte, Concordia University, Canada – North American indigenous art

### **Competing Interests**

The authors have the following competing interests to declare:

Waqas Ahmed is Artistic Director of the Khalili Collections

Martin Poulter is the Wikimedian in Residence at the Khalili Collections

### **Contributions**

#### ***Authorial contributions***

Authorship is alphabetical after the drafting author and principal technical lead. Author contributions, described using the CASRAI CredIT typology, are as follows:

Author name and initials:

Waqās Ahmed (WA)

Martin Lewis Poulter (MP)

Authors are listed in descending order in each category by significance of contribution:

Corresponding author: WA

Conceptualization: WA

Data Curation: MP

Formal Analysis: MP  
 Funding Acquisition: WA  
 Investigation: WA, MP  
 Methodology: MP  
 Project Administration: WA, MP  
 Software: MP  
 Validation: MP  
 Visualization: WA, MP  
 Writing – Original Draft Preparation: WA, MP  
 Writing – Review & Editing: WA, MP

### **Editorial contributions**

#### **Section Editor**

Morgan Pearce, The Journal Incubator, University of Lethbridge, Canada

#### **Copy Editors**

Morgan Pearce, The Journal Incubator, University of Lethbridge, Canada

Akm Iftekhhar Khalid, The Journal Incubator, University of Lethbridge, Canada

Christa Avram, The Journal Incubator, University of Lethbridge, Canada

#### **Layout Editor and Production Consultant**

Virgil Grandfield, The Journal Incubator, University of Lethbridge, Canada

### **References**

- Callahan, Ewa S., and Susan C. Herring. 2011. "Cultural Bias in Wikipedia Content on Famous Persons." *Journal of the American Society for Information Science and Technology* 62(10): 1899–1915. DOI: <https://doi.org/10.1002/asi.21577>.
- Dean, Carolyn. 2006. "The Trouble with (the Term) Art." *Art Journal* 65(2): 24–32. DOI: <https://doi.org/10.2307/20068464>.
- Dolmaya, Julie McDonough. 2017. "Expanding the Sum of All Human Knowledge: Wikipedia, Translation, and Linguistic Justice." *The Translator* 23(2): 143–157. DOI: <https://doi.org/10.1080/13556509.2017.1321519>.
- Graham, Mark, and Matthew Zook. 2013. "Augmented Realities and Uneven Geographies: Exploring the Geolinguistic Contours of the Web." *Environment and Planning A: Economy and Space* 45(1): 77–99. DOI: <https://doi.org/10.1068/a44674>.

- Graham, Mark, Scott A. Hale, and Monica Stephens. 2011. *Geographies of the World's Knowledge*, edited by Corinne M. Flick. London: University of Oxford. Oxford Internet Institute. Convoco! Edition. Accessed July 18, 2022. <https://ora.ox.ac.uk/objects/uuid:a737b5c3-069f-4044-ade8-2ce6c95a0d4c>.
- Halfaker, Aaron. 2017. "Interpolating Quality Dynamics in Wikipedia and Demonstrating the Keilana Effect." *OpenSym '17: Proceedings of the 13<sup>th</sup> International Symposium on Open Collaboration* 19: 1–9. DOI: <https://doi.org/10.1145/3125433.3125475>.
- Maher, Katherine, and Loic Tallon. 2018. "Wikimedia and The Met: A Shared Digital Vision." *Metropolitan Museum of Art*, April 19. Accessed June 27, 2022. <https://www.metmuseum.org/blogs/now-at-the-met/2018/wikimedia-and-the-met-digital-vision>.
- Maurer, Hermann, and Josef Kolbitsch. 2006. "The Transformation of the Web: How Emerging Communities Shape the Information We Consume." *Journal of Universal Computer Science* 12(2): 187–213. Accessed September 7, 2022. [https://www.jucs.org/jucs\\_12\\_2/the\\_transformation\\_of\\_the.html](https://www.jucs.org/jucs_12_2/the_transformation_of_the.html).
- MediaWiki Contributors. 2021. "API: Main Page." MediaWiki. Accessed January 24, 2021. [https://www.mediawiki.org/wiki/API:Main\\_page](https://www.mediawiki.org/wiki/API:Main_page).
- Meta Contributors. 2021a. "Wikimedia Movement 2017 Strategy/ Direction." Meta-Wiki. Accessed January 24, 2021. [https://meta.wikimedia.org/wiki/Strategy/Wikimedia\\_movement/2017/Direction](https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2017/Direction).
- . 2021b. "Wikimedia Movement Strategy Recommendations." Meta-Wiki. Accessed January 4, 2021. [https://meta.wikimedia.org/wiki/Strategy/Wikimedia\\_movement/2018-20/Recommendations/Identify\\_Topics\\_for\\_Impact](https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2018-20/Recommendations/Identify_Topics_for_Impact).
- . 2021c. "Wikimedian in Residence." Meta-Wiki. Accessed January 24, 2021. [https://meta.wikimedia.org/wiki/Wikimedian\\_in\\_residence](https://meta.wikimedia.org/wiki/Wikimedian_in_residence).
- OpenGLAM Working Group. 2011. "OpenGLAM Principles, Version 1.0." Accessed January 24, 2021. <https://openglam.org/principles/>.
- Orlowitz, Jake. 2018. "The Wikipedia Library." In *Leveraging Wikipedia: Connecting Communities of Knowledge*, edited by Merrilee Proffitt, 69–86. Chicago: American Library Association.
- Outreach Wiki Contributors. 2021a. "GLAM/Repository." Outreach Wiki. Accessed January 4, 2021. <https://outreach.wikimedia.org/wiki/GLAM/Repository>.
- . 2021b. "GLAM-WIKI." Outreach Wiki. Accessed January 24, 2021. <https://outreach.wikimedia.org/wiki/GLAM>.
- Steiner, Christopher B. 1996. "Can the Canon Burst?" *Art Bulletin* 78(2): 213–217.
- Wikidata Contributors. 2022. "Kawade Shibataro." Reasonator. Accessed April 29, 2022. <https://reasonator.toolforge.org/?&q=21664940>.
- Wikimedia Commons Contributors. 2021. "Commons: Guide to Content Partnerships." Wikimedia Commons. Accessed January 4, 2021. [https://commons.wikimedia.org/wiki/Commons:Guide\\_to\\_content\\_partnerships](https://commons.wikimedia.org/wiki/Commons:Guide_to_content_partnerships).



Wikipedia Contributors. 2020. "Wikipedia: Vital Articles." Wikipedia. Accessed November 1, 2021. [https://en.wikipedia.org/wiki/Wikipedia:Vital\\_articles](https://en.wikipedia.org/wiki/Wikipedia:Vital_articles).

———. 2021a. "Help: Adding Open License Text to Wikipedia." Wikipedia. Accessed January 4, 2021. [https://en.wikipedia.org/wiki/Help:Adding\\_open\\_license\\_text\\_to\\_Wikipedia](https://en.wikipedia.org/wiki/Help:Adding_open_license_text_to_Wikipedia).

———. 2021b. "Wikipedia: WikiProject Women in Red." Wikipedia. Accessed January 4, 2021. [https://en.wikipedia.org/wiki/Wikipedia:WikiProject\\_Women\\_in\\_Red](https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Women_in_Red).

### Further readings

Beach, Milo Cleveland. 2011. *Masters of Indian Painting, 1100–1900*. Zürich: Verl. Museum Rietberg.

Dalrymple, William, Lucian Harris, Rosie Llewellyn-Jones, J. P. Losty, H. J. Noltie, Malini Roy, Yuthika Sharma, and Andrew Topsfield. 2019. *Forgotten Masters: Indian Painting for the East India Company*. London: Philip Wilson Publishers.

Goswamy, Brijinder Nath. 2016. *The Spirit of Indian Painting: Close Encounters with 101 Great Works, 1100–1900*. London: Allen Lane.

Guy, John, and Jorrit Britschgi. 2011. *Wonder of the Age: Master Painters of India, 1100–1900*. New York: Metropolitan Museum of Art.

O'Brien, Elaine, Everlyn Nicodemus, Melissa Chiu, Benjamin Genocchio, Mary K. Coffey, and Roberto Tejada. 2012. *Modern Art in Africa, Asia, and Latin America: An Introduction to Global Modernisms*. Chichester, UK: Wiley-Blackwell.

