

Image-based information: paintings in Wikipedia

By Trilce Navarrete and Elena Villaespesa

Abstract

Purpose – This study aimed at understanding the use of paintings outside of an art-related context, in the English version of Wikipedia.

Method – For this investigation, we identified 8,104 paintings used in 10,008 articles of the English Wikipedia edition. We manually coded the topic of the article in question, documented the number of monthly average views, and identified the originating museum. We analysed the use of images based on frequency of use, frequency of view, associated topics, and location. Early in the analysis three distinct perspectives emerged: the readers of the online encyclopaedia, the editors of the articles, and the museum organisations providing the painting images (directly or indirectly).

Findings – Wikipedia is a widely used online information resource where images of paintings serve as visual reference to illustrate articles, notably also beyond an art-related topic and where no alternative image is available – as in the case of historic portraits. Editors used paintings as illustration of the work itself or art-related movement, but also as illustration of past events, as alternative to photographs, as well as to represent a concept or technique. Images have been used to illustrate up to 76 articles, evidencing the polysemic nature of paintings. We conclude that images of paintings are highly valuable information sources, also beyond an art related context. We also find that Wikipedia is an important dissemination channel for museum collections. While art related articles contain greater number of paintings, these receive less views than non-art related articles containing fewer paintings. Readers of all topics, predominantly history, science and geographic articles, viewed art pieces outside of an art context. Painting images in Wikipedia receive a much larger online audience than the physical painting does when compared to the number of museum onsite visitors. Our results confirm the presence of a strong long-tail pattern in the frequency of image use (only 3% of painting images are used in a Wikipedia article), image view, and museums represented, characteristic of network dynamics of the Internet.

Research limitations/Implications – While this is the first analysis of the complete collection of paintings in the English Wikipedia, our results are conservative as many paintings are not identified as

such in Wikidata, used for automatic harvesting. Tools to analyse image view specifically are not yet available and user privacy is highly protected, limiting the disaggregation of user data. This study serves to document a lack of diversity in image availability for global online consumption, favouring well-known Western objects. At the same time, the study evidences the need to diversify the use of images to reflect a more global perspective, particularly where paintings are used to represent concepts of techniques.

Practical implications – Museums wanting to increase visibility can target the reuse of their collections in non-art related articles, which received 88% of all views in our sample. Given the few museums collaborating with the Wikimedia Foundation and the apparent inefficiency resulting from leaving the use of paintings as illustration to the crowd, as only 3% of painting images are used, suggests further collaborative efforts to reposition museum content may be beneficial.

Social implications – This paper highlights the reach of Wikipedia as information source, where museum content can be positioned to reach a greater user group beyond the usual museum visitor, in turn increasing visual and digital literacy.

Originality/Value – This is the first study that documents the frequency of use and views, the topical use and the originating institution of ‘all the paintings’ in the English Wikipedia edition.

Keywords – Museums, Paintings, Wikipedia, Image-based information

1. Introduction

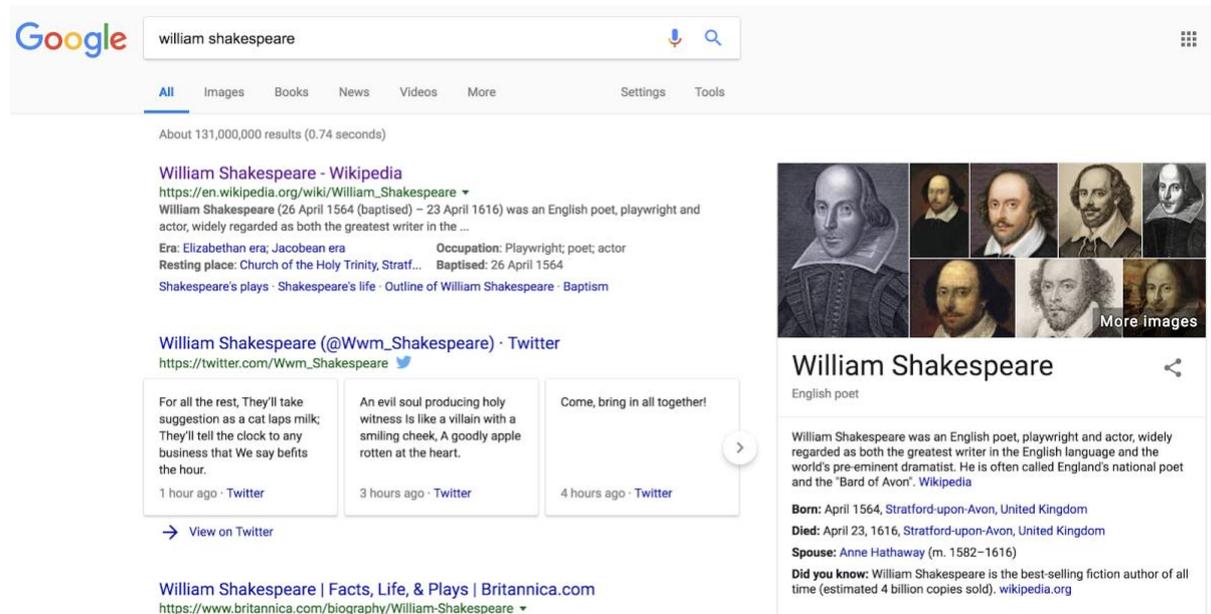
We live in an information economy, where information assumes a central role in every aspect of life, where new producers and new information networks create new services to provide new ways of accessing and using information (for an earlier discussion see Saracevic, 1997:528). In our increasingly visual culture, image literacy is gaining relevance, as understanding compositional elements enables interpretation and communication based on the meaning inferred (Lopatovska, et al., 2016; Marsh and Domas, 2003).

A specific type of image-rich information is found in heritage collections housed in galleries, libraries, archives, and museums (GLAMs). Perhaps two of the best-known examples of new providers that have developed content-rich information services free-riding on the millenary collections are Google Books or the Google Arts and Culture Institute. Google has identified the undeniable great information value of heritage collections to advance its information services.

Another relatively new information service is Wikipedia, with a remarkably high-ranking score. The user journey to get to a Wikipedia page starts in 65% of the cases with a search in Google (Alexa ranking),¹ showing links to article content at the top of the results page, often also in the knowledge graph box next to the available images. After landing on broad articles, consumers will navigate to more specific content (Rodi et al., 2017). Once in an article, images play an important role for readers to assess trustworthiness, as images contribute positively to the appearance and value of the article, following textual features and references (Lucassen and Schraagen, 2010). This is particularly so for individuals lacking previous knowledge on the topic (Lucassen, et al., 2013).

¹ <https://www.alexa.com/siteinfo/wikipedia.org>.

Figure 1. Query result for 'William Shakespeare' including painting from the National Portrait Gallery in London



The free online encyclopaedia was launched in 2001 and remains one of the top 5 most visited websites worldwide, being an important provider of information for humans and for machines. While there is a noticeable amount of research on the online encyclopaedia (a Google Scholar query results in over 2 million entries while a Web of Science query results in over 5,000 results), there is limited understanding about the use of images (Lucassen and Schraagen, 2010; Lucassen, et al., 2013) and currently no overview of the use of GLAM images to illustrate articles. Images from museums' collections are increasingly being used to illustrate articles and only a noticeable small number of museum institutions have collaborated to support the creation of articles, about their institution, their collections, or a variety of topics (Stinson, Fauconnier and Wyatt, 2018).

Considering that museums' purpose revolves around providing access to collections, in the present and in the future, the online publication of collection information to increase access would appear as a desired option. Digital access benefits from reduced restrictions of time and space as well as cost of entry, for the most part. From the last European survey on digital access to heritage collections, 21% of respondents reported viewing cultural heritage-related content online while 50% reported physically visiting a museum or gallery (increasing from 37% reported in 2013) (EU, 2017). Museums have a history of presenting their collections online as information, building from a greater tradition of managing information about the objects in catalogues, displays, for access and for interpretation (Burton Jones, 2008). Few have ventured to publish collections online in non-art related contexts.

In this paper, we look at Wikipedia as an information exchange centre (an information technology) and we focus on collections information from museums all over the world (as image-based information). Our goal is to understand the use of image-based heritage information particularly outside of an art-related context, while documenting the polysemic information value of collections.

From an information science perspective, the value of the information system is defined by the user. Taking the core elements of information science as posited in the seminar work of Saracevic (1999, 2004, 2007a, 2007b, 2008), we explore the application of his framework to study the use of the image-based information goods.

The remainder of the paper is organised as follows: section two presents key concepts of information science, with focus on relevance, followed by the use of images as information. Section three presents our data collection method and section four presents the results. In section five we discuss the application of the framework based on our data analysis. We close with conclusions and draw lines of future research in section six.

2. Related literature

2.1 The information science perspective

Information science typically studies the “effective communication of knowledge records among humans in the context of social, organizational, and individual need for and use of information” (Saracevic, 1999:1056). Research can be categorized in one of the three main contributions of the field: on information retrieval (the processing of information), on relevance (effectiveness relative to a human), or on interaction (exchange between humans and information systems). The value of information and information services is closely related to relevance, defined as “a *dynamic* concept that depends on users’ judgments of the quality of the relationship between information and information need at a certain point in time” (Schamber, Eisenberg, and Nila, 1990:771, italics by the authors).

While the study of information utility in economic analysis has focused on ratings and reviews as quality signals to lower the information asymmetry problem, information scientists have studied the interaction between systems and actors seeking information to evaluate relevance. The manifestations of relevance are: (1) system relevance, measured as query result dependent on text retrieved as processed by an algorithm, based on comparative effectiveness; (2) subject relevance,

measured as topic processing or queried about the record ('aboutness');² (3) cognitive relevance, measured as the correspondence between user, record, and system, defined by cognitive correspondence, informativeness, novelty, and information quality; (4) situational relevance, measured as the utility of a record for a specific situation, dependent on usefulness; and (5) motivational relevance, defined by the user satisfaction, success and accomplishment of the intent of the query (Saracevic, 1999). The latter is referred to as socio-cognitive relevance by Cosijn and Ingwersen, (2000) who propose motivation to be an attribute of relevance (as part of intention) instead of a manifestation. Attributes of relevance are defined as the value dimensions of the information and include relation, intention, context, inference, and interaction (Saracevic, 1997a).

Relevance is closely linked to time, it increases with greater cognitive effect and decreases with greater effort to process information (White, 2017). In other words, relevance is dependent on expertise, the greater the domain knowledge the greater the inference of relevance (Vakkari and Hakala, 2000 in Saracevic, 2007a; White, 2017). It is to be expected that an art historian, compared to a non-specialist, is able to identify a greater number of relevant articles where a painting may serve as illustration, or gain additional knowledge from the information in a painting, as previous knowledge informs cognitive associations. Further, the evaluation of images as part of a document has been associated with the information skills that enable an individual to judge information, which gain importance for domains outside of the individual's area of expertise (Lucassen, et al., 2013). Images can thus influence relevance of a document by facilitating information processing.

Relevance can be created or derived through inference (Saracevic, 2007a), and so images of paintings can be made relevant to illustrate a Wikipedia article, for instance, while receiving relevance from the content of the article. Creating and deriving relevance takes place in a continuum, in which users select the most appropriate knowledge record in a dynamic process of interaction, interpretation, and evaluation. The most relevant document will be the most effective in serving to carry the information sought, again dependent on the specific context. This continuum is defined by the socio-cognitive context of the user (Cosijn and Ingwersen, 2000).

2.2 Images as information

The use images as information representation of a museum object is one of many other principles of representation (e.g. catalogue card records) that assist management and reuse of information housed in museum collections historically. The advantages of using surrogates, or information representation,

² Fujita (2000) proposes a 'foreground' subject and a 'background' subject to information objects, where the foreground subject may be the name of the individual portrayed in a painting and the background subject the landscape to position the scene, for example. This is a useful distinction on aboutness as representation of information objects since all paintings contain layered subjects.

are associated with information transmission at great distances, with interactivity between users and objects, and with being able to target unique information needs (Marty, 2008). Milekic (2007) argues information transfer of abstract information is more efficient through the use of tangibles, such as images. Technological advances have allowed non-text-based information to become predominant in the information rendering of dynamic and immersive experiences (Cameron and Robinson, 2007).

Images are perceived as supplementary to textual information but are, in fact, crucial information references and quality signals. Consumers do appreciate relevant, accurate images (Choi and Rasmussen, 2001), even if they do not necessarily see images as information sources. This seemingly contradiction is further evident in that museums are perceived as trusted traditional repositories of images, ranked as the second highest trusted source of information following libraries, but are not a common used source of information (Usherwood, et al. 2005). On the other hand, easily accessible information sources such as television, radio, tabloids, and increasingly the Internet were used the most though trusted the least.

Given the polysemic nature of museum collections, it is surprising that these are not used more prominently outside of the museum setting. It is with each temporary exhibition that museums reinterpret objects in new contexts. Cameron and Robinson (2007) have argued that the given meaning of an object is not self-evident but imposed by each museum institution, precisely by the classification and descriptive categorization to manage information.

2.3 Incentives to use images as information in Wikipedia

Venture capital for image platforms may be available for profitable businesses and a few museums make available their images for licensing. Wikipedia, as non-profit information platform with education and knowledge transfer spearheading its activities, is a natural partner for museums (and all GLAMs) in the information market. Wikipedia increases its information quality by associating itself with established content providers of diverse, quality, and organised heritage information while museums benefit by tapping into the global network. Museums and other heritage institutions have attempted to create a global platform for heritage content (Europeana.eu), yet this is hardly known by the general public. An important shortcoming to image online publication is copyright, yet this is not expected to be the major inhibitor since less than 15% of collections have licensing rights by a third party (Nauta, et al., 2017).

While Wikipedia does not provide funds to digitise collections, they do organise the crowd to photograph objects exhibited physically in museums and to use existing images of museum collections online as article illustrations, lowering the labour costs for museums. The Wikimedia Foundation further provides a number of services that may turn attractive to museums institutions, including the

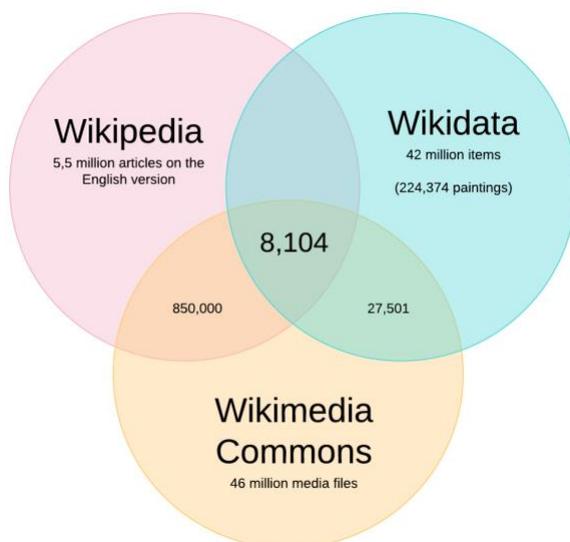
Wikimedia Commons image repository and the WikiData structure, all with a respectable history: Wikipedia was launched in 2001 and continues to grow, which cannot be said of many museum websites. Wikipedia regularly ranks among the 10 top websites, currently ranking number 5 according to Alexa, has over 5 million articles in English, and includes articles in over 250 languages. A second key project is Wikimedia Commons, which was launched in 2004 and serves as repository of images, sounds, videos and general media. As its name suggests, all 40 million media files are freely available with an open creative commons license (CC0, CC BY, and CC BY-SA). Over 850,000 media files are used in the English edition of Wikipedia. One last project key for our data collection is Wikidata, launched in 2012 as a common source of data for items used in all projects. Wikidata has over 42.3 million data items feeding Wikipedia and other projects, over 5 million for the English Wikipedia alone.

Analysis of free consumption of heritage image online is limited to harmonised available data, which excludes Google's Books and Art Project. We therefore chose the next global alternative that while relatively embryonic still serves to identify important patterns of heritage image use online. Our analysis serves to explore the use of paintings outside an art context, as illustrations in Wikipedia.

3. Data

Data was gathered from three main projects of the Wikimedia Foundation: Wikidata, Wikimedia Commons, and Wikipedia. Figure 2 illustrates the relation between the three Wikimedia Foundation projects and our dataset.

Figure 2. Relation of Wikimedia projects and dataset (June 2017)



A SPARQL query was conducted on 13 June 2017 from the Wikidata endpoint. We started by identifying the ‘paintings’ available in Wikidata, resulting in 224,374 items. Table 1 shows the overview of the data, including items labelled as paintings, containing basic metadata (author, date of creation, and location), that have an image representation, and that are used in a Wikipedia article in the English edition. Noticeable is that less than half of ‘paintings’ have basic identifying information, that less than a fourth have an image, and that only 4% of all ‘paintings’ in Wikidata are used in a Wikipedia article. For a detailed description of the SPARQL query see Annex A1. We chose to triangulate the data from three Wikimedia projects for analysis because there is no API for accessing all information about the images from the Commons repository. Instead, Wikidata and Wikipedia both provide a downloadable dataset that can be queried using SPARQL.

Table 1. Summary of paintings in Wikidata

Wikidata set	Items count	Percentage
Paintings	224.374	100
With metadata	89.637	40
With image	27.501	12
In Wikipedia	10.008	4
In Wikipedia articles	8,104	3
Museums (or collection)	785	-
Country location	59	-

3.1 Coding

We manually assigned a code following the frequently used category ontology proposed by Spoerri (2007) in order of importance: Entertainment, Politics and History, Geography, Sexuality, Science, Computers, Arts, Religion, Holidays, Current events, and Drugs (see Table A2 for categories and sub-categories). During the manual coding, paintings were assigned a sub-category. Deviating from Spoerri, Sports was separated from Entertainment, and History was divided from (current) Politics. These sub-categories were kept due to their frequency of use and to highlight the actual use of museum paintings, notably to illustrate sports and to differentiate historic articles from politics related articles. We added a category of ‘Wikipedia’ to exclude pages that are not properly an encyclopedic article such as ‘did you know’ and ‘features’ of articles or images, file names, templates, and lists (e.g. paintings, years, recent additions). This left a dataset of 8,104 paintings (3% of paintings in Wikidata) that were used in 10,008 Wikipedia articles (some used multiple times).

Using a Python script that connected to the Wikipedia pageviews API (Application Programming Interface), we harvested the monthly page views from the period January through to December 2017 in all identified Wikipedia articles (N=10,008). The official definition of a page view is “a request for the content of a web page” (Wikimedia: Research: Page view). It is worth noting that page view does not equal unique users, therefore, a user (reader) can view multiple Wikipedia articles and every time a new article is loaded during the timeframe is counted as a page view. Table 2 shows the number of paintings identified by category and the percentage of Wikipedia articles it represents, with the monthly average views per page.

Table 2. Summary of harvested dataset

Categories	Number of articles	%	Article views	%	Number of paintings	%
Art (paintings)	5,516	68.1%	11,239,450	11.9%	3,250	32.5%
History	3,034	37.4%	41,037,732	43.6%	4,354	43.5%
Religion	924	11.4%	10,172,624	10.8%	897	9.0%
Geography	600	7.4%	13,739,610	14.6%	589	5.9%
Other art forms (fashion, performing arts, literature)	533	6.6%	3,306,198	3.5%	314	3.1%
Science	447	5.5%	10,923,875	11.6%	450	4.5%
Entertainment	60	0.7%	1,313,478	1.4%	75	0.7%
Sexuality	50	0.6%	1,606,074	1.7%	29	0.3%
Sport	26	0.3%	616,753	0.7%	40	0.4%
Politics	11	0.1%	116,517	0.1%	12	0.1%
Drugs	2	0.0%	196,261	0.2%	2	0.0%
Grand Total	8,104		94,217,895		10,008	

Adding to the harvested dataset, we manually sought the museum source of the paintings represented in our study, whenever possible as some location points were not always clear (e.g. ‘storage space’ or ‘private collection’). We identified a total of 785 museums (or collection) located in 59 countries. In addition, we manually identified the number of yearly physical visitors of the museums in the sample as indicator of institutional size. Information was gathered from the institutional website or the annual reports.

3.2 Limitations

Before continuing onto the analysis of our dataset, a few limitations are worth mentioning. The first and most important constraint to our study is the identification of ‘paintings’ in Wikipedia articles. Our choice of using Wikidata to identify an instance of a painting allows for mechanical analysis,

prohibitively labour intensive otherwise, though excludes paintings lacking a Wikidata profile. Hence images not identified as paintings in Wikidata are not included in our dataset. This is mostly the case of paintings added by individual editors, from books or various sources, rather than by the museum institution. A manual check on the page of Alexander Hamilton reveals the presence of 12 paintings (and a total of 29 media files) of which our dataset only includes one. Similarly, our dataset excludes the second most visited museum of the world, the National Museum of China. Our results appear to fall on the conservative side. Future research could also include other art forms, such as sculpture and photography. We further excluded paintings lacking an image and basic metadata (author, date of creation).

Secondly, we are looking only at the English version of Wikipedia, which will reflect a usage pattern that may not be shared in other language editions. A recent study (Singer, et al., 2017) of the users of Wikipedia reflects indeed that each language edition has characteristic dynamics. Looking at the dynamics of using paintings as illustrations in multiple languages editions of Wikipedia promises a rich line of future work.

Third, the manual classification of the categories of the paintings inevitably reflects personal subjectivity in the process. For instance, we have classified all the definitions of actions, objects, and customs as Science (honouring humanities), specific places and cultural customs as Geography (e.g. Spanish food). For a detailed list of topics included in each category, see Table A2. The main goal of the classification exercise was to identify the Wikipedia articles that are directly related to the arts in general, including the painting, the creator, the technique, the art movement or the exhibiting museum, as oppose to those articles using paintings within seemingly unrelated topics (e.g. science, sports). Other categorisation schemes may facilitate a different analysis, for instance by looking at the position of the painting within the article, or the topic in the painting used (e.g. portrait, abstract).

Last, our dataset is a snap shot in time allowing us to make observations of a current situation present in the English Wikipedia mid 2017 based on what there is. We are limited to 'views' and cannot identify 'viewers', which would highlight multiple page view or repeat visits, or individuals performing several tasks at different times, such as a museum editing content. Even if we are unable to establish casual relations we can, however, suggest relations based on previous related literature as well as additional datasets. Ideally, our results will provide a stepping stone for future surveys of editors, readers, and museums to better understand relevance of paintings, and further museum collections, as illustrations of an online encyclopaedia.

4. Results

We examined the use of paintings as illustrations of the English Wikipedia and in the process distinguish three distinct user groups: (1) the editors of Wikipedia articles, responsible for selecting the image that will illustrate the article at hand; (2) the readers, seeking information; and (3) the museums, responsible from selecting images to be made digitally available from their vast collections and serving as main providers of the images analysed. The data available is aggregated and does not allow to distinguish unique users on Wikipedia. However, it is worth noting that one individual may play one or more roles though generally individuals have one role at one moment in time. In the following section, we discuss the observed behaviour of the three user groups in relation to the images of paintings from museums forming our dataset.

4.1 Editors

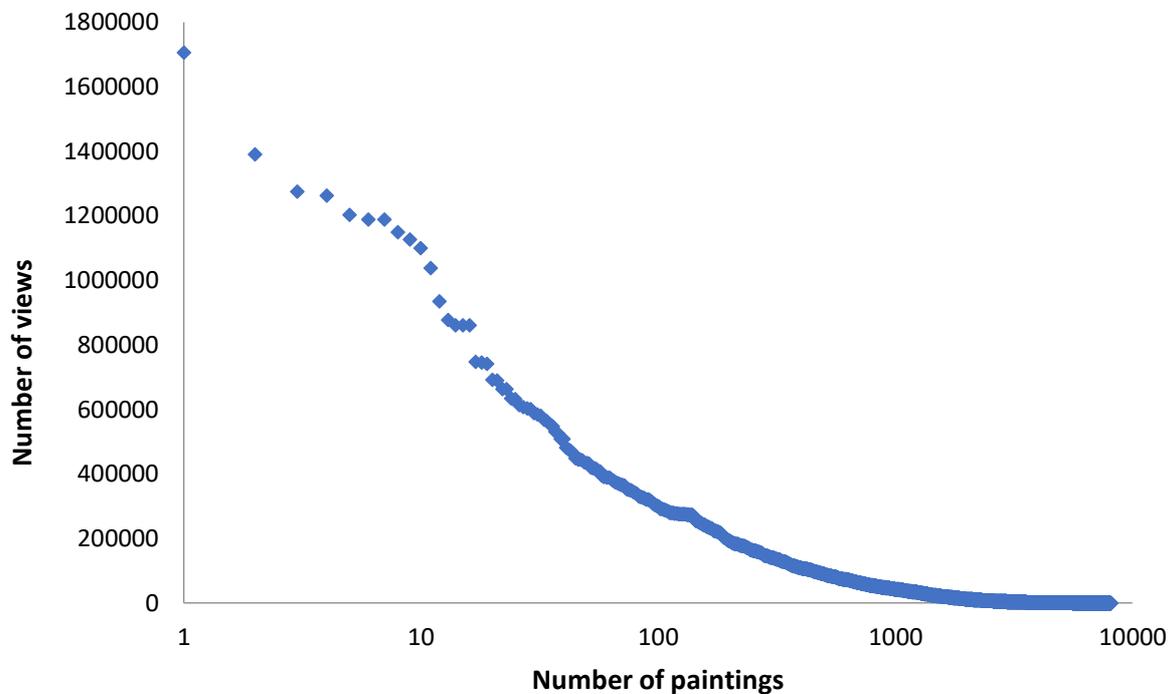
This analysis of the first group looks at how painting images have been added to Wikipedia articles. From the 27,500 images of paintings in our dataset (with 'basic' metadata), we identified 8,104 paintings (30%) are included in 10,008 unique articles. There are paintings included in more than one article and sometimes multiple images from the dataset are found in the same article (e.g. Van Gogh article includes 33 images from the dataset).

The distribution of the usage of paintings on Wikipedia articles presents a long-tail, where one item is used in 163 Wikipedia pages, two items in 76 pages, the following 90 items are used in 20 to 50 pages, followed by close to 1,000 items used in less than 20 pages, while 5,351 painting images were used in just one article (Figure 3). The most used painting is a portrait of William Shakespeare by John Taylor from 1610 and the one that gets most views is *Mona Lisa* by Leonardo da Vinci from 1503 (See top 20 most used painting images on Table 3). Surprising was the use of Shakespeare's portrait to illustrate a 'Collar' (in clothing) while the *Mona Lisa* is found as example of the colour 'Green'. Several colours have their own article that is illustrated by a number of images, including paintings.

Though it can be expected that paintings illustrate art, heritage, and history related articles, we find geography, science, and sexuality also using paintings as illustrations. Not surprisingly, art related pages have more paintings per page in comparison with other article categories. Another interesting insight is that one artwork image can be used to illustrate a range of topics. For instance, the painting 'The Scream' by Edvard Munch is highlighted in the article about the painter as well as in the art-related articles covering the art movements 'Expressionism' and 'Modern art'. However, the articles using Munch's painting as illustration that get the most views are 'Anxiety disorder', 'Meaning of Life' and 'Krakatoa'. The last article about the volcano in Indonesia includes a description of the eruption

of the volcano in 1883, which reached Norway the year in which the painting was made, explaining the colour choice of Munch for the sky in the painting.

Figure 3. Wikidata ‘paintings’ views on Wikipedia (logarithmic scale) (N=8,104 paintings)



We find that paintings are used in Wikipedia as illustration in the following three main modalities.

(1) Paintings illustrate the work itself, and an article is made to inform about the maker, the artistic technique, the historical context or further artistic context. Wikipedia articles about museums often list the top paintings in their collections.

(2) Paintings serve to illustrate the past, as alternative to photography, for portraits, battles, locations, extinct animals, and when other images are unavailable. An example is the article about Spain which includes several paintings that illustrate the country’s history (e.g. Portrait of Philip II, El Tres de Mayo, Philip V’s family). Due to the type of paintings found in the National Portrait Gallery in the United Kingdom, the collection appears in numerous articles about political and historical figures (e.g. Charles I of England, George II of Great Britain) but also scientists and engineers (e.g. James Watt), writers (e.g. Charles Dickens), or composers (e.g. George Frideric Handel).

(3) Paintings are also found to support the illustration of symbols, concepts, and techniques. Examples include the articles of ‘Liberty’ with Eugene Delacroix’s ‘Liberty leading the people’ from 1830, the ‘Starry Night’ by Van Gogh symbolising ‘Psychosis,’ a Portrait of Luca Pacioli by Jacopo de Barbari

(1495) in ‘Accounting,’ the painting of a Girl by Juan Carreno de Miranda (1680) in ‘Obesity,’ or the portrait of Percy Bysshe Shelley by Alfred Clint (1819) in ‘Black.’ Paintings are particularly useful to illustrate difficult topics, as the case of Titian’s Tarquin and Lucretia in the article ‘Rape,’ or the Anatomy Lesson of Dr. Tulp by Rembrandt to illustrate an ‘Autopsy’.

Table 3. Top 20 most used paintings illustrating a Wikipedia article

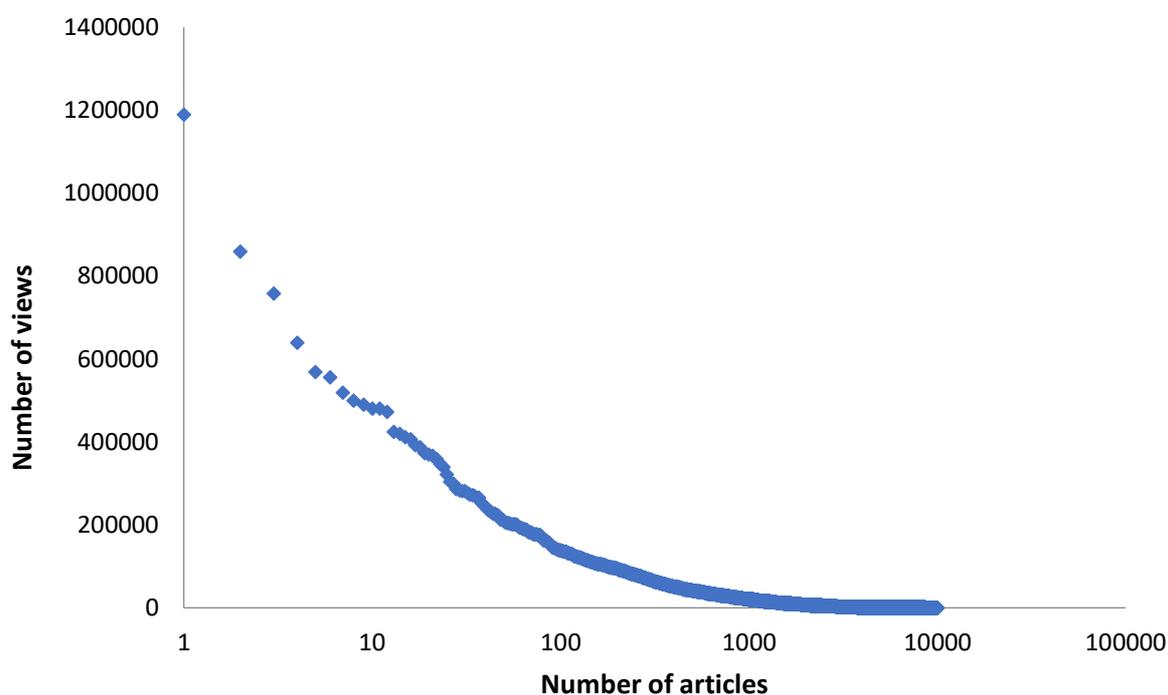
N.	Painting image	Number of articles	Total views (monthly)
1	William Shakespeare by John Taylor	76	1,263,648
2	Liberty Leading the People by Eugène Delacroix	44	1,099,838
3	Bacchus by Peter Paul Rubens	37	663,508
4	Mona Lisa by Leonardo da Vinci	35	1,707,053
5	Surrender of General Burgoyne by John Trumbull	34	548,878
6	The Princes in the Tower by John Everett Millais	32	410,082
7	Surrender of Lord Cornwallis by John Trumbull	31	1,203,959
8	Portrait of Louis XIV (1638-1715), King of France by Hyacinthe Rigaud	27	747,439
9	Washington Crossing the Delaware by Emanuel Leutze	26	606,791
10	Sir Walter Scott by Henry Raeburn	24	691,926
11	Robert Burns, 1759 - 1796. Poet by Alexander Nasmyth	23	195,885
12	Portrait of Erasmus by Hans Holbein	22	877,508
13	The Birth of Venus by Sandro Botticelli	22	663,819
14	The Arnolfini Portrait by Jan van Eyck	22	328,507
15	Portrait of Philip II by Sofonisba Anguissola	21	745,975
16	Sir Thomas More by Hans Holbein	20	278,590
17	Portrait of Henry VIII of England by Hans Holbein	20	1,390,269
18	Samuel Johnson by Joshua Reynolds	20	304,820
19	Scene at the Signing of the Constitution of the United States by Howard Chandler Christy	20	1,128,131
20	The Scream by Edvard Munch	20	635,481

Selection process and motivation behind the choice of the one image over other by editors falls outside of the scope of this paper. Further image analysis to determine relevance based on style of images chosen to illustrate articles may drive a future research line.

4.2 Readers

The second user group analysis examines the readership of the Wikipedia articles that include these images. As expected due to the popularity of some articles in the sample, the chart of the distribution of the number of views shows a sharp long-tail shape (see Figure 4).

Figure 4. Wikipedia article views (logarithmic scale) (N=10,008)



The same finding about the distribution of article views applies to the distribution of views within each of the article categories. As it can be observed in Table 4, the standard deviation is high which is due, for the majority of categories, to the degree of information seeking interest of the topics covered in the articles. Articles related to science and geography, that include a painting image as illustration, receive significant more views than visual art related articles.

Table 4. Statistics of article views per category

	No. of articles	Unique	Average	Median	std. dev.
Art	5,516	11,434,992	3,478	549	13,126
History	3,034	41,154,249	9,426	885	37,928
Religion	924	10,172,624	11,341	1,403	24,342
Geography	600	13,739,610	23,327	1,938	69,029
Science	447	10,923,875	24,275	12,407	49,090
Literature	112	2,077,892	12,906	2,025	43,697
Performing arts	130	1,032,764	8,981	812	26,160
Entertainment	60	1,313,478	17,513	2,219	43,854
Sport	26	616,753	15,419	1,658	35,518
Sexuality	50	1,606,074	55,382	26,241	77,719

Drugs	2	196,261	98,130	98,130	92,815
-------	---	---------	--------	--------	--------

Table 5 presents the list of the 20 most viewed articles, all pages of countries and names of well-known historical figures. We see that several articles have more than one painting and that many articles are not related to art. It must be noted that we can only speak of views to articles, and hence paintings, and not on viewers of the articles, or individuals viewing multiple articles. Nevertheless, the skewed visibility of certain paintings, based on the article these are positioned at, is striking.

Table 5. Top 20 most viewed Wikipedia articles containing a painting

N.	Article	Category	Views (monthly)	Number of paintings
1	Queen Victoria	History	1,189,122	2
2	Charles Darwin	Science	859,127	1
3	Canada	Geography	758,040	1
4	Mary, Queen of Scots	History	639,455	1
5	Abraham Lincoln	History	568,828	2
6	Russia	Geography	555,588	2
7	France	Geography	518,598	2
8	Alexander Hamilton	History	499,887	2
9	Henry VIII of England	History	490,664	2
10	American Civil War	History	481,034	1
11	William Shakespeare	Literature	480,899	1
12	Elizabeth I of England	History	471,910	1
13	Indonesia	Geography	424,255	1
14	George V	History	420,157	1
15	George Washington	History	411,002	3
16	Netherlands	Geography	405,950	3
17	Spain	Geography	392,483	6
18	England	Geography	386,848	1
19	Rothschild family	History	372,806	1
20	Christopher Columbus	History	369,951	1

Isolating the art-related pages, Table 6 shows the number of paintings used in each page is higher than in pages of non-art related topics. The most viewed pages are about well-known artists, art pieces, and art periods.

Table 6. Top 20 most viewed art-related Wikipedia pages including a painting

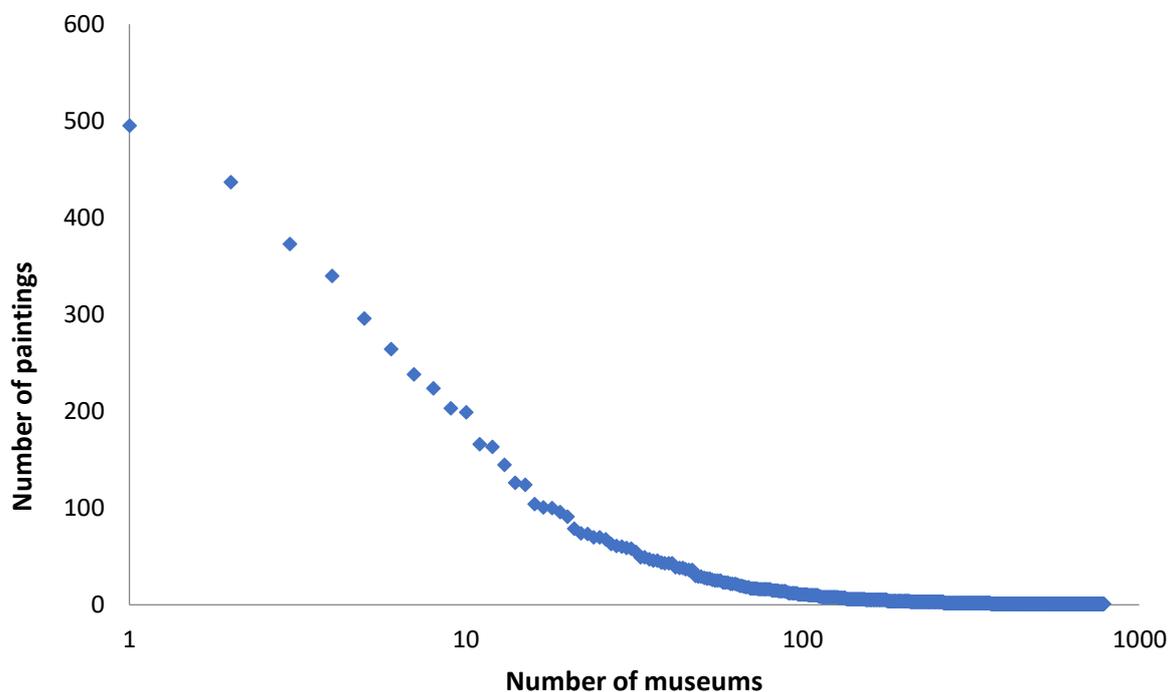
N.	Article	Views (monthly)	Number of paintings
1	Leonardo da Vinci	359,195	5
2	Vincent van Gogh	272,892	33
3	Mona Lisa	265,626	3
4	Michelangelo	162,605	2
5	Romanticism	135,136	21
6	Art	105,672	2
7	The Starry Night	97,641	3
8	David (Michelangelo)	93,108	1
9	Claude Monet	90,716	20
10	Louvre	90,678	8
11	Impressionism	90,573	17
12	The Scream	89,131	3
13	A Midsummer Night's Dream	84,488	2
14	William Blake	83,126	2
15	Modernism	82,459	1
16	Rembrandt	73,509	28
17	Raphael	72,205	14
18	Georgia O'Keeffe	67,820	2
19	Guernica (Picasso)	66,707	1
20	Art Nouveau	66,345	3

On average, article readership remains relatively steady across the year, with a ten percent fluctuation peaking in January, May, and October, suggesting a close relation to academic cycles. A closer look to the path followed by the reader, particularly when navigating through painting's hyperlinks, remains for the future research agenda.

4.3 Museums

The third user group analysis expands the harvested dataset with data on annual visitation available on museum websites and financial reports. Annual physical visits serve as indicator of institutional size in order to compare institutions. It is expected that larger number of visits can be found in institutions with larger collections and greater budgets. This additional information led to an interesting comparison between physical visitors viewing the collections and Wikipedia views. We identified 785 museums that have at least one painting illustrating a Wikipedia article. 54% have only one painting in Wikipedia while 2%, have more than one hundred paintings used as illustrations in Wikipedia articles (See Figure 5). As in the two previous user group analysis, we find a long-tail.

Figure 5. Museum collections represented in Wikipedia (logarithmic scale) (N=785)



When we analyse the number of views, the biggest museums in the world are at the top of the list (see Table 7). These museums have large collections made up of well-known artists and artworks and are a must visit in major cities, hence known as superstar museums (Frey, 2000). Paintings of mid-size museums receive a greater number of views when used to illustrate typical encyclopaedic articles. That is the case of the National Portrait Gallery in London, which provides portrait illustrations to dozens of articles about historical figures such as Charles Darwin, William Shakespeare, Charles Dickens, James VI and I, and Charles I of England.

Other popular pages include the article 'Sleep' which includes the painting 'The Land of Cockaigne' by Pieter Bruegel (1567) from the Alte Pinakothek in Germany, receiving eighty thousand monthly views, or the article 'Black' which includes six paintings from our dataset and receives 54 thousand monthly views. *The portrait of Portrait of Philip II of Spain* by Sofonisba Anguissola in the Museo del Prado's collection in Madrid and the *Portrait of a Young Woman* by Petrus Christus from the Gemäldegalerie in Berlin collection exemplify how the colour black was used in Europe in the 15th and 16th centuries.

Table 7. Top 20 most prominent museums by number of article views

N.	Museum / Collection	Views	Articles	Paintings	Onsite visitors
1	Louvre Museum	10,846,301	933	495	8,100,000
2	National Portrait Gallery (London)	8,184,376	766	373	1,703,411
3	National Gallery	8,048,944	728	340	5,229,192
4	Museo del Prado	7,192,095	481	238	2,824,404
5	Metropolitan Museum of Art	6,316,214	516	296	7,000,000
6	National Gallery of Art	4,342,791	411	224	5,232,000
7	Musée d'Orsay	4,275,493	303	203	3,178,000
8	Hermitage Museum	3,855,089	327	199	4,200,000
9	Rijksmuseum	3,614,756	522	437	2,148,304
10	Tate	3,481,855	336	163	8,400,000
11	United States Capitol	3,473,796	91	4	3,000,000
12	Kunsthistorisches Museum	2,175,747	270	264	559,150
13	Walters Art Museum	2,041,227	118	70	NA
14	National Galleries of Scotland	1,967,830	121	43	2,200,000
15	Galerie des Batailles	1,904,632	124	28	NA
16	Thyssen-Bornemisza Museum	1,871,305	80	49	1,052,014
17	Uffizi	1,863,024	147	68	2,010,631
18	Tretyakov Gallery	1,846,755	113	43	2,024,000
19	Buckingham Palace	1,825,699	29	14	NA
20	"Museum's storage space"	1,810,379	217	166	NA

It is noticeable that except for the Louvre and the Musée d'Orsay, receiving nearly the same number of physical visits as online views to Wikipedia articles containing their images, all other institutions receive a significant larger number of views to a much smaller number of paintings via the online encyclopaedia. The collections of the Alte Pinakothek in Berlin receive more than four times the amount of views online than onsite, followed by the National Portrait Gallery in London receiving nearly four times more views, and the Kunsthistorisches museum and Boijmans van Beuningen in Rotterdam receiving nearly three times more views. On the other hand, there are museums with large number of physical visits that receive little online visibility, such as the Vatican Museum in Rome with only 5 paintings in Wikipedia articles or the Centre Pompidou in Paris with only 1 painting.

Inclusion in Wikipedia articles requires images have a CCO license, with variants including BY or SA. This legal requirement may explain the limited amount of museum collections available in Wikimedia Commons, particularly of the contemporary collections, as museums struggle to clear legal permission to enable further use of images.

The lack of diversity in museum representation in our dataset leaves much room for future research, including the inclusion of other museum objects (beyond painting) which may be more representative of global collections, for instance including photographs and heritage objects, but also other language editions (eg. French, Chinese, or Spanish Wikipedia).

5. Discussion

Wikipedia, and partner projects, is an important global information platform. We have identified three distinct user perspectives from which to analyse the use of paintings outside of a museum, in this case as images illustrating the online English Wikipedia. We can first of all confidently state that the information platform has proven to be relevant worldwide for many years (for our three user groups as well as for machines), given the constant growth of articles and views since it was launched. When considering the specific user groups, it is clear that editors have an important role in the selection of paintings to illustrate images as they define the topical context within which a painting will be used in a Wikipedia article. Their selection will define the relevance experienced by readers of the image as well as the article (relevance created through inference), and future editors repositioning images. Editors first identified images of museum collections, including paintings, as relevant image references, before museum institutions became involved. Little is known about Wikipedia editors regarding their process of editing and reason for selecting paintings, as opposed to other images or other media, or their awareness that certain images originate from museum collections. Based on the images used, and the articles chosen to be illustrated, an editor favours clear metadata to identify the image (such as 'portrait of' or with title 'solitude' to illustrate such entry) than abstract or subjective topics. An interesting case is the use of Laura Knight's 'Ruby Loftus screwing a breech ring' painting from 1943, housed at the Imperial War Museum in London, to illustrate the article of 'Occupational safety and health' (see Figure 6). Museums may want to consider collaborating with the editors to expand the number of images used, which may require greater specialised knowledge.

How do editors find paintings? A list of paintings as category is not available for content in the repository (Wikimedia Commons), though change is underway,³ and knowledgeable editors may know about the lists of paintings per year or the Sum of All Paintings, a Wikidata project to include all 'notable' paintings.⁴ Given the prominent role of Wikipedia to feed the Internet, greater diversity of image use in Wikipedia articles (and increasingly also on the Wikidata structure), can be expected to

³ A project is underway to structure the data in Commons, based on Wikidata, see https://commons.wikimedia.org/wiki/Commons:Structured_data.

⁴ https://www.wikidata.org/wiki/Wikidata:WikiProject_sum_of_all_paintings.

trickle down to other online domains. The complexity found in cultural content evidences the limited choices of paintings by editors.

Figure 6. Painting illustrating a non-art related article



The image shows a screenshot of a Wikipedia article page. At the top left is the Wikipedia logo and navigation links. The article title is "Occupational safety and health". Below the title is a sub-header "From Wikipedia, the free encyclopedia". The main text discusses the field of Occupational Safety and Health (OSH), its goals, and its legal implications. On the right side, there is an image of a woman working on a lathe in a factory, with a caption explaining that she is not wearing eye protection, which would be illegal in modern industrialized countries.

Readers are perhaps the best-known users of Wikipedia. A recent study (Singer, et al. 2017) identifies heterogeneous behaviour across language editions and three main uses of Wikipedia: for fact checking, in-depth reading, or for overview of topics. The report found a significant correlation between socio-economic indicators and Wikipedia use, where higher GDP countries use the mobile version for fact checking while lower GDP countries use the desktop version for in-depth reading. The report did not look into image use. Our data shows images are used when no alternative image is available, particularly to the non-art related topics. This is the case of The Muir portrait serving as portrait of 'Adam Smith', housed at the Scottish National Gallery (see Figure 7). It can be expected that the portrait contributes to process information about the 18th century, including the customary use of wigs, for readers not familiar with the knowledge domain (Lucassen, et al., 2013). Image paintings of botanicals or (extinct) animals may be important contributors of information transfer (Milekic, 2007; Choi and Rasmussen, 2001).

Policy implications call for a greater representation of paintings that are not only made by the Western superstar painters but that also include other views of the world. Illustrations that reference

information recognizable in diverse contexts can be expected to facilitate information transfer for a greater number of individuals (Cosijn and Ingwersen, 2000). We expect the use of paintings in Wikipedia articles will contribute to an increase in the general visual capital of users, particularly on non-art related topics, making diversity of image use critical. Encountering images clearly labelled as resource from a museum may further raise awareness of museums as valuable information sources, with reusable content, solving the perceived discrepancy of trust versus use identified by Usherwood et al., (2005).

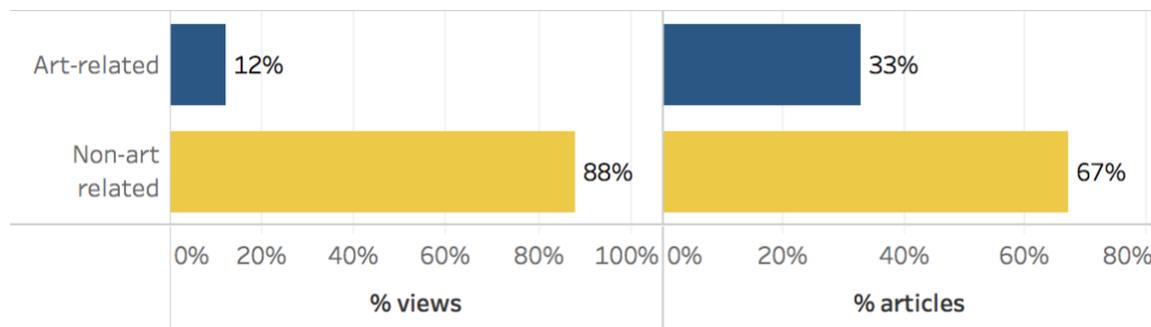
Figure 7. Painting illustrating a notable figure

The image shows a screenshot of the Wikipedia article for Adam Smith. At the top right, it says "Not logged in" with links for "Talk", "Contributions", "Create account", and "Log in". Below this is a search bar and navigation tabs for "Article", "Talk", "Read", "Edit", and "View history". The article title is "Adam Smith" with a lock icon. Below the title, it says "From Wikipedia, the free encyclopedia". A note reads: "For other people named Adam Smith, see Adam Smith (disambiguation)." The main text describes Adam Smith as a Scottish economist, philosopher, and author, a pioneer of political economy, and a key figure during the Scottish Enlightenment. It mentions his works, *The Theory of Moral Sentiments* (1759) and *An Inquiry into the Nature and Causes of the Wealth of Nations* (1776). A portrait painting of Adam Smith is shown on the right, with a caption "Adam Smith FRSA". Below the portrait is a table of his biographical information:

Born	16 June [O.S. 5 June] 1723 Kirkcaldy, Fife, Scotland
Died	17 July 1790 (aged 67) Edinburgh, Scotland
Nationality	Scottish
Alma mater	University of Glasgow Balliol College, Oxford
Notable work	<i>The Wealth of Nations</i> <i>The Theory of Moral Sentiments</i>

The most interesting result to emerge from the data is that paintings illustrating non-art related articles receive 88% of views from all articles in our sample. Figure 8 shows the striking share of views to articles not directly linked to the topic of art. This result highlights the information value of museum collections, particularly beyond an art context, and urges museums to exploit the potential information use of their collections.

Figure 8. Share of art and non-art related views to Wikipedia articles containing a painting



Regarding our third user group, museums, we observe Wikipedia facilitates the use of paintings as image-based information in an unprecedented way, from museums from all over the world. Considering the polysemic nature of collections (Cameron and Robinson, 2007), articles can be found about any possible topic, just like exhibitions displaying paintings from museum collections. Editors may not be familiar with the rich collections in each museum or with the historic system to classify and give meaning to objects used by museums (Marty, 2008). Similarly, readers may not associate museum collections with relevant information sources, yet value the visual reference when readily available (Choi and Rasmussen, 2001). Museums appear to slowly adapt Wikipedia as valuable active partner to continue presentation and dissemination of collections as information (Burton Jones, 2008; see Villaespesa and Navarrete, 2019, for an overview of museums collaborating with Wikipedia).

We observe by the use of ‘categories’ and file naming of images that a number of paintings are donated by museums in collaboration with Wikipedia. However, we also see many paintings are ‘found’ by editors to fill an illustration gap, either through the Google Art and Culture Institute (former Google Art Project) or as book scans. An example can be found in the article ‘Russia’ including a painting by Ivan Shishking depicting Rye, housed at the Tretyakov Gallery and taken from the Google Art Project (see Figure 9).

The fact that the Louvre Museum has been the most visited museum in the world for several years, and its paintings in Wikipedia receive the greatest number of views, is clear evidence of the power of the superstar. For all non-superstar museums, the number of views in the online platform is of a different dimension when compared to the number of physical visits (Navarrete and Borowiecki, 2016). Google Art and Culture Institute has expanded its reach to increase representation of lesser known museums. Future analysis to track changes in the use of images in Wikipedia beyond the superstars, may provide insights in the dynamics of image reuse on the Internet.

Figure 9. Painting taken from the Google Art Project

Agriculture

Main articles: [Agriculture in Russia](#) and [Fishing industry in Russia](#)

Russia's [total area of cultivated land](#) is estimated at 1,237,294 square kilometres (477,722 sq mi), the fourth largest in the world.^[244] From 1999 to 2009, Russia's agriculture grew steadily,^[245] and the country turned from a grain importer to the third largest grain exporter after the EU and the United States.^[246] The production of meat has grown from 6,813,000 tonnes in 1999 to 9,331,000 tonnes in 2008, and continues to grow.^[247]

This restoration of agriculture was supported by a credit policy of the government, helping both individual farmers and large privatized corporate farms that once were Soviet [kolkhozes](#) and which still own the significant share of agricultural land.^[248] While large farms concentrate mainly on grain production and [husbandry](#) products, small private [household plots](#) produce most of the country's potatoes, vegetables and fruits.^[249]

Since Russia borders three oceans (the Atlantic, Arctic, and Pacific), Russian fishing fleets are a major [world fish supplier](#). Russia captured 3,191,068 tons of fish in 2005.^[250] Both exports and imports of fish and sea products grew significantly in recent years, reaching \$2,415 and \$2,036 million, respectively, in 2008.^[251]

Sprawling from the Baltic Sea to the Pacific Ocean, Russia has more than a fifth of the world's forests, which makes it the largest forest country in the world.^{[182][252]} However, according to a 2012 study by the [Food and Agriculture Organization of the United Nations](#) and the Government of the Russian Federation,^[253] the considerable potential of Russian forests is underutilized and Russia's share of the global trade in forest products is less than four percent.^[254]



Rye Fields, by Ivan Shishkin.

Russia is the world's top producer of [barley](#), [buckwheat](#) and [oats](#), and one of the largest producers and exporters of [rye](#), [sunflower seed](#) and [wheat](#).

Given the coarse data available for the analysis, we are limited in what we can say about Saracevic's information science framework on the processing of information (information retrieval), on the effectiveness relative to a human (relevance) and on the exchange between humans and information systems (interaction). From our data analysis we observed the retrieval of information by editors is partially evident in the selection of paintings to illustrate an article, favouring literal illustrations and reuse of images over greater diversity of selection and abstract representation. Further causes, incentives, and actual process requires a different research. For readers, we observed that involuntary retrieval of images of paintings is prevalent over voluntary art-related searches, as the majority of views are reported for non-art related articles. Regarding museums, we identified 785 museums to be linked to at least one Wikipedia article. Understanding the institutional process to make illustrations available for retrieval can drive a future investigation. A further proposal on how analyse the retrieval, interaction, and relevance of paintings used as illustrations in Wikipedia articles can be found in Table A3 in the annex.

6. Conclusions

Museums are rich repositories of image-based information that can be viewed online. We have investigated Wikipedia as information system where paintings serve to illustrate art and non-art related articles. We identified all paintings used in the English Wikipedia articles (containing the label

'painting' in Wikidata), manually coded the topic of the article, documented the readership frequency as number of monthly average views, and manually linked each painting to a museum institution. We identified three distinct user groups: editors, readers, and museum institutions. The rest of the analysis took one of the three user perspectives.

We find a sharp disparity in the articles receiving more views, on the paintings most used, and on the representation of museums. The most viewed articles correspond to historic, geographic and religious subjects, mostly receiving under a half million monthly views, with 'Queen Victoria' (over one million monthly views) and 'Charles Darwin' (just under one million monthly views) topping the list. The majority of paintings are used in less than 10 articles, where the most used paintings are a portrait of William Shakespeare (76 articles) followed by *Liberty Leading the People* (44 articles) and *Bacchus* (37 articles). The majority of museums represented have less than one hundred paintings illustrating articles, with the Louvre (495 paintings) and the Rijksmuseum (437 paintings) being the most present collections. The popularity of certain paintings, articles, and museums may be inevitable within a large information network. Noticeable is the potential increase in online views with relatively small amount of collection objects.

Upon more detailed inspection, we identified that paintings are often used as literal representations based on their title or image metadata (e.g. portrait of individual) facilitating image retrieval by editors. Given that complex and abstract tasks are more time consuming and require greater human capital, it is no surprise that most paintings are used as direct literal illustrations (e.g. historical portrait) rather than for secondary topics (e.g. 'Occupational safety and healthy'). This finding supports previous evidence of the positive effects of using standardize metadata systems to catalogue museum collections. 'Untitled' paintings are less likely to be easily found by editors.

We find that the majority of views are received by paintings in non-art related articles. This is an important finding, as it suggests that paintings are highly relevant information goods often relegated within a purely art context. Museums can increase active participating in the information network by enabling images to serve a much greater and broader user group, which is visible when comparing Wikipedia visibility to number of physical visitors.

Museums are rich repositories of knowledge records that can gain relevance by being accessible beyond a specific subject or location. Wikipedia provides an environment where each painting can be positioned within multiple articles, serving multiple information purposes at the same time, for a global audience. Paintings are polysemic information goods, which gain relevance by serving as illustration for articles while making articles relevant by providing additional information. Our results

serve as starting point to further understand the potential market for heritage visual content in the information economy.

This is the first time, as time of writing as far as the authors are concerned, that the entire corpus of paintings is analysed in the English Wikipedia version. Major limitations include the availability of data, which depended on Wikidata to flag 'paintings' from the media repository, and the current organisation of topics in Wikipedia. Future research is welcomed to compare other language editions, to include other museum objects (such as sculptures, photographs, and crafts), and to expand the scope of variables analysed.

References

- Briet, Suzanne (1951) *Qu'est-ce que la documentation*. Paris: EDIT.
- Buckland, Michael (1997) 'What is a Document' in *Journal of the American Society for Information Science* (1986-1998). 48(9):804-809.
- Burton Jones, Katherine (2008) 'The Transformation of the Digital Museum' in Marty, Paul and Katherine Burton Jones (Eds.) *Museum Informatics. People, Information, and Technology in Museums*. New York: Routledge. Pp.9-25.
- Cosijn, Erica and Peter Ingwersen (2000) 'Dimensions of relevance' in *Information Processing and Management*. 36(4):533-550.
- Choi, Youngok and Edie Rasmussen (2001) 'Users' relevance criteria in image retrieval in American history' in *Information processing and management*. 38(5):695-762.
- EU (European Union) (2017) Special Eurobarometer 466, Cultural Heritage. Wave EB88.1, DG COMM.
- Frey, Bruno (2000) *La Economia del Arte*. Barcelona: La Caixa.
- Fujita, Sumio (2001) 'Reflections on aboutness TREC-9 evaluation experiments at just system' in *Proceedings of the ninth text retrieval conference (TREC-2001)*. Gaithersburg, MD:NIST Special Publication, pp.281-288.
- Hamid, Rahayu, Thom, James and Iskandar, Awang (2016) 'Effects of relevance criteria and subjective factors on web image searching behaviour' in *Journal of Information Science*. 43(6):786-800.
- Lopatovska, Irene, Hatoum, Sarah, Waterstraut, Saebra, Novak, Lisa, and Sheer, Sara (2016) 'Not just a pretty picture: visual literacy education through art for young children' in *Journal of Documentation*. 72(6):1197-1227.
- Lucassen, Teun, and Schraagen, Jan (2010) 'Trust in Wikipedia: How Users Trust Information from an Unknown Source' in *WICOW'10*, 27 April 2010, Raleigh, NC, USA.
- Lucassen, Teun, Rienco Muilwijk, Matthijs Noordzij, and Jan Maarten Schraagen (2013) 'Topic Familiarity and Information Skills in Online Credibility Evaluation' in *Journal of the American Society for Information Science and Technology*. 64(2):254:264.
- Marsh, Emily and Marilyn Domas White (2003) 'A taxonomy of relationships between images and text' in *Journal of Documentation*. 59(6):647-672.

- Marty, Paul (2008) 'Information Representation' in Marty, Paul and Katherine Burton Jones (Eds.) *Museum Informatics. People, Information, and Technology in Museums*. New York: Routledge. Pp.29-34.
- Nauta, Gerhard Jan, Van den Heuvel, Wietske and Teunisse, Stephanie (2017) *Europeana DSI 2 – Access to Digital Resources of European Heritage*. D4.4. Report on ENUMERATE Core Survey 4. The Hague: DEN Foundation.
- Navarrete, Trilce and Borowiecki, Karol (2016) 'Change in access to heritage after digitisation: ethnographic collections in Wikipedia' in *Cultural Trends*, 25(4):233-248.
- Otlet, Paul (1934) *Traité de documentation*. Brussels: Editiones Mundaneum.
- Rodi, Giovanna, Loreto, V, and Tria, F. (2017) 'Search strategies of Wikipedia readers' in *PLOS one*. 12(2): e0170746. DOI:10.1371/journal.pone.0170746.
- Saracevic, Tefko (2007a) 'Relevance: A Review of the Literature and a Framework for Thinking on the Notion in Information Science. Part II: Nature and Manifestations of Relevance' in *Journal of the American Society for Information Science and Technology*. 58(13):1915-1933, 2007.
- Saracevic, Tefko (2007b) 'Relevance: A Review of the Literature and a Framework for Thinking on the Notion in Information Science. Part III: Behavior and Effects of Relevance' in *Journal of the American Society for Information Science and Technology*. 58(13):2126-2144, 2007.
- Saracevic, Tefko (2008) 'Effects of Inconsistent Relevance Judgments on Information Retrieval Test Results: A Historical Perspective' in *Library Trends*.56(4):763-783.
- Saracevic, Tefko (2004) 'Evaluation of digital libraries: an overview' in Agosti, M. and Fuhr, N. (eds.) *Notes of the DELOS WP7 Workshop on the evaluation of Digital Libraries*, Padua, Italy. Available at: http://dlib.ionio.gr/wp7/workshop2004_program.html.
- Saracevic, Tefko (1999) 'Information Science' in *Journal of the American Society for Information Science*.
- Saracevic, Tefko and Paul Kantor (1997) 'Studying the Value of Library and Information Services. Part 1. Establishing a Theoretical Framework' in *Journal of the American Society for Information Science*. 48(6):527-542.
- Schamber, Linda, Michael Eisenberg, and Michael Nilan (1990) 'A re-examination of relevance: toward a dynamic, situational definition' in *Information Processing and Management*. 26(6):755-776.
- Singer, Philipp, et al. (2017) 'Why We Read Wikipedia' in *Proceedings of the 26th International Conference on World Wide Web, WWW 2017*, pp.1591-16---. DOI 10.1145/3038912.3052716.

- Spoerri, Anselm (2007) 'What is Popular in Wikipedia and Why' in *First Monday*. 12(4):1-15.
- Stinson, Alexander, Fauconnier, Sandra, and Wyatt, Liem (2018) 'Stepping Beyond Libraries: The Changing Orientation in Global GLAM-Wiki' in *JLIS.it* 9,3 (September 2018):16-34. DOI: [10.4403/jlis.it-12480](https://doi.org/10.4403/jlis.it-12480).
- Usherwood, Bob, Kerry Wilson, and Jared Bryson (2005) 'Relevant repositories of public knowledge? Libraries, museums and archives in "the information age"' in *Journal of Librarianship and Information Science*. 37(2):89-98.
- Villaespesa, Elena, and Trilce Navarrete (2019) 'Museum Collections on Wikipedia: Opening Up to Open Data Initiatives'. *MW19: Museums and the Web 2019*. Consulted February 3, 2019. <https://mw19.mwconf.org/paper/museum-collections-on-wikipedia-opening-up-to-open-data-initiatives/>
- White, Howard (2017) 'Relevance theory and distributions of judgements in document retrieval' in *College of Computing and Informatics*. 53(5):1080-1102.
- Wikimedia: Research: Page view. Consulted May 24, 2019. https://meta.wikimedia.org/wiki/Research:Page_view

Annex

Annex A1. SPARQL query

On 13 June 2017, we conducted a SPARQL query from the endpoint at <https://query.wikidata.org>:

```
SELECT ?item ?itemLabel ?loc ?locLabel ?inc ?cre ?creLabel
WHERE
{
  ?item wdt:P31 wd:Q3305213 ;
    wdt:P276 ?loc ;
    wdt:P571 ?inc ;
    wdt:P170 ?cre .
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en" }
}
```

This query returns all items in the Wikidata database that are *an instance of (wdt:P31) a painting (wd:Q3305213)*:

- the item ID and verbal description thereof
- its location ID (wdt:P276) and verbal description thereof
- its inception date (wdt:P571)
- its creator ID (wdt:P170) and verbal description thereof

The results were saved as .csv file. The intention was to identify the items labelled as 'painting' from Wikidata in order to identify the items used in Wikipedia articles.

The Wikidata file was cleaned as follows: from the 224,374 paintings found in Wikidata, we selected those items including basic metadata of location, date of creation and author. 2,045 duplicates were identified, as some paintings were attributed to different makers resulting in a double entry of the same painting. That left a dataset of 89,637 unique items labelled as 'painting' with basic metadata. We then selected the paintings that included and image resulting in 27,501 unique images. Last, we counted the number of instances these images were mentioned on a Wikipedia page, resulting in 10,054 paintings.

Table A2. Categories of Wikipedia articles (based on Spoerri, 2007)

Primary category	Secondary category
Entertainment	Comics, films, games, musical, pop music, sports, TV series, video games;
History and politics	Battles, historic documents, elections, events, explorers, political figures, historical figures, law, historical periods, social movements, wars;
Geography	Countries, culture, places;
Sexuality	Sexual body parts, related activities;
Science	Formal science, natural science, and social science: Anatomy, archaeology, astronomy, biology, chemistry, earth sciences, humanities, law, linguistics, medicine, scientists, species;
Computers	
Arts (Paintings)	Art, artists, museums, paintings;
Other art forms	Books, dance, fashion, literature, music, performing arts and artists, playwrights, writers;
Religion	Mythology, religious depictions, religious terms;
Holidays	
Current events	
Drugs	Substances, practices;
Wikipedia	Book:, category:, featured article, featured picture, file, lists, goings-on, main page, manual, peer review, portal, recent additions, signpost, template, Wikipedia:, WikiProject.

It was decided to adjust the original Spoerri (2007) categories to better fit the categories encountered. Though Entertainment contained all music and books, we included historical music (e.g. virginal, opera) and literature (e.g. King Lear) in Art and kept musicals and pop music (e.g. album, CD) in Entertainment. Geography includes cultural groups (e.g. Slovene literature) as well as specific locations (e.g. Notre Dame de Paris). Sexuality includes articles related to definitions (e.g. rape, lesbian) and cultural practice (e.g. nude swimming). Science includes all formal, natural and social sciences, including colours, linguistics, sociology, and law. Arts includes artists' biographies, historic fashion, and most art work, excluding religious depictions (e.g. Adoration of the Magi) which were included in Religion. Art subcategories were identified for museums and for fashion. Drugs included substances (absinthe) and practices (cannabis culture). The Wikipedia category included all the pages organized by the association, including features, lists, special pages and reports. There were no article allocations to Computers, to Holidays, and to Current events categories.

Table A3. Framework to study relevance of paintings as visual references in Wikipedia (proposed)

	Editors	Readers	Museums
Retrieval	Selection of paintings to illustrate article. Responsible for establishing contextual information (caption, article)	Search for information, use of information (not necessarily images).	Providers of images (to be used as illustrations and to be viewed in articles).
Interaction	Editing of content within the Wikimedia platforms (Wikidata, Commons, Wikipedia).	Reading Wikipedia through browser or mobile app.	Uploading content into the Wikimedia platforms.
Relevance	Dependent on finding the 'right' image for the article at hand.	Dependent on ability to gain additional information through an image.	Contingent to desirability of system to distribute digital images.
System relevance (text retrieved)	Dependent on category and metadata of image.	Dependent on article.	Provider of metadata, keywords, and image title.
Subject relevance (topic queried)	Establish relation between image and subject in article.	Image can provide and/or gain subject relevance in article.	All objects in the collection are potentially relevant for at least one article.
Cognitive relevance (user, record, system)	Established based on own cognitive ability, levels of subtlety reflected on use of one painting in multiple articles, or use of obscure paintings.	Depending on cultural capital (and knowledge of paintings) to understand various complexity levels of references.	Portraits are popular, paintings used as alternative to photography.
Situational relevance (usability of record)	May be dependent on findability within Wikimedia Commons.	May be dependent on prominence of image, or on number of hyperlinks in article. May be less relevant for mobile factual check.	Improved with relevant keywords and image title.
Motivational relevance (user satisfaction)	Paintings as alternative for photographs, as symbol, as illustration, as contextual information.	Unless specifically seeking the painting, access of record is not intentional. Satisfaction may vary.	Dependent on motivation for open data publication. All images used in articles report greater views.