THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Measuring bias in aggregated digitised content

Citation for published version:
Kizhner, I, Terras, M, Rumyantsev, M, Khokhlova, V & Demeshkova, E 2019, Measuring bias in aggregated digitised content: A case study on Google arts and culture. in *Digital Humanities 2019, Utrecht .* Alliance of Digital Humanities Organisations. <https://dev.clariah.nl/files/dh2019/boa/0467.html>

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Digital Humanities 2019, Utrecht

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 21. May. 2022

DH2019
UTRECHT UNIVERSITY 8-12 JULY

Utrecht University

Measuring Bias in Aggregated Digitised Content: a Case Study on Google Arts and Culture

XML

Large cultural heritage aggregators, such as Europeana and Google Arts and Culture (GA&C), collect metadata and images from cultural institutions. They provide a single portal that introduces cultural heritage from around the world to the public (Sood, 2016, Petras et al., 2017). Selecting images and artifacts for these aggregators is an outcome of curatorial decisions, enlarging an art canon (Earhart, 2012, Feldman, 2016), building a cultural capital (Bertrand and Kamenica, 2018), and providing an infrastructure for a corpus of art history images (Drucker, 2013) that is critically important for the research in Digital Humanities. However, are such portals indeed a representative and balanced collection, the foundation for objective humanistic study and judgement? In this paper we argue that diversity, although present in GA&C, is too narrow to support our hope that it can act as a corpus of digital art history images. Our evidence proves that the digital corpus amplifies biases within the arts world towards western culture.

1. Methodology

1.1. The source of GA&C collections

Our analysis is based on the full collection of two-dimensional images with metadata available on GA&C web site (ca. 5,000,000 images), excluding videos and street view panoramas. All our numeric estimates below are based on *Collections* (museum collections or collections of other holders, such as LIFE magazine or Opera national de Paris). We accessed collections via *Places* option in the GA&C menu and identified the size of the collections submitted to GA&C and the country of the collection holders.

1.2. Russian collections in GA&C

GA&C's collections from Russia are relatively small (ca. 5000 images) supplied by 49 institutions. We have identified geographical location of all 49 GA&C contributors from the Russian Federation and for the 32 from them that are listed in the official museum registry of the Russian Ministry of Culture and are currently tabulating genres, periods and authorship of the 2,802 images that these museums supplied to GA&C. This is a limited share of the artifacts held by over 2,000 Russian museums that altogether hold about 60,000,000 objects in the main part of their collections.

1.3. French collections in GA&C

We identified 87 French collections that we used as a control group and compared them with the list of over 1,200 French museums downloaded from the web site of the French Ministry of Culture. We found 19 museums from the list of the Ministry with 4,477 images of objects.

We understand that two countries can be hardly enough to find out how representative GAC is and further research is needed to cover a larger sample of countries.

2. Results

2.1. Is Google Arts and Culture a balanced corpus?

Western museum collections have traditionally been biased in approaching art objects in the frames of western

aesthetic/cultural concepts (Chalmers, 1996, Ang, 2005, Simpson, 2012) and the bias continues in new digital aggregators. Table 1 shows that the majority of images come from the top five countries at the time of writing this paper. It demonstrates that a vast majority of providing cultural institutions come from the United States, United Kingdom and the Netherlands. The largest collection (Table 2) is the LIFE Photo Collection, New York, with 4,5 million images (80% of all images). The fifteen countries to follow the top five are represented with a much smaller number of objects. However, all the countries from the list of the United Nations are represented in the aggregator either as countries with institutional collections or countries tagged as the places of origin of cultural objects (*Discover this place* group). Our analysis shows that 123 countries out of 195 nations are not represented in the aggregator through their collections but can be referred to as the places of origin.

Table 1

The number of records with images for the institutional collections representing the places related to the names of countries in GAC. The images from the top five countries' collections account for 93% of images

Country	**No of images, institutional collections**	% of total
USA	4,713,779	82 %
United Kingdom	334,558	5.8
Netherlands	170,201	3 %
Italy	98,225	1.7 %
South Korea	52,214	0.9 %
Other countries (190 countries from the UN list)	337,198	6.6 %

Table 2

The images from the top four collections account for 88% of images

Collections	Country, city	**Number of images, institutional collections**	% of total
LIFE Photo Collection	USA, New York	4,403,372	79.3
The Natural History Museum	UK, London	298,804	5.4
Rijksmuseum	Netherlands, Amsterdam	164,510	3
The Strong National Museum of Play	USA, Rochester	72,556	1.3
Other collections	Other	615,113	11,1

2.2. Do smaller collections represent provincial museums to demonstrate diversity?

Our results demonstrate that although the aggregator can be considered a representative corpus at the scale of nations, it is by no means a balanced corpus. We show that provincial museums take a small portion of the museums represented by the aggregator for the two countries in the study.

The paper will provide the distribution of genres, periods, geographical regions and artists when the study is completed to be presented at Digital Humanities 2019.

3. Conclusion

We demonstrate that GA&C is a representative collection of images as it includes at least two images from the cultural institutions of every country recognized by the United Nations. However, our results indicate that five countries (UK, USA, Netherlands, Italy, South Korea) contributed the largest share of images. Our hypothesis is that GA&C tends to work with the museums that are either easier to work with due to common legislation structures, or similar attitudes and languages, those museums that are more accessible in terms of opening their collections or with private museums that are interested in making their collections known. Further research is needed to provide evidence in support of this assumption.

Introducing cultural heritage that was previously difficult to access seems to be the task that new platforms perform fairly well. However, the case of GA&C demonstrates that this digital corpus tends to reinforce traditional biases of western curatorship (Manovich, 2017). This results in important consequences for large scale research in Digital Humanities when obtained patterns are skewed towards western aesthetic and cultural concepts. It may also prevent the general public from building a cultural capital based on cultural diversity.

Appendix A

Bibliography

1. **Ang, I.** (2005). The predicament of diversity: multiculturalism in practice at the art museum. *Ethnicities,* 5(3): 305-320.
2. **Bertrand, M. and E. Kamenica.** (2018). Coming Apart? Cultural Distances in the United States over Time. NBER working paper no. 24771, http://faculty.chicagobooth.edu/marianne.bertrand/research/papers/comingApartOnline.pdf (accessed 25 November 2018).
3. **Chalmers F.G.** (1996). *Celebrating pluralism: art, education and cultural diversity*. Los Angeles *:* The Getty Educational Institute for the Arts.
4. **Drucker J.** (2013), Is there a 'digital' art history? *Visual Resources*, 29(1–2): 5–13.
5. **Earhart A.** (2012). Can information be unfettered? Race and the new digital humanities canon. In *Debates in the Digital Humanities*. Minneapolis, MN: University of Minnesota, pp. 309–18.
6. **Feldman, M.H.** (2016). Rethinking the Canon of Near Eastern Art in the Internet Age, *Journal of Ancient Near Eastern History,* 3: 57-79.
7. **Grady, J.** (2007). Advertising images as social indicators: depictions of blacks in LIFE magazine, 1936-2000. *Visual Studies,* 22(3): 211-239.
8. **Manovich, L.** (2017). Cultural Data: Possibilities and Limitations of Working with Historical Cultural Data. In Oliver Grau (Editor), Museum and Archive on the Move: Changing Cultural Institutions in the Digital Era (pp. 259–276). Berlin, Boston: De Gruyter.
9. **Petras, V., Hill, T., Stiller, J. and Gäde, M.** (2017), Europeana – a search engine for digitized cultural heritage material, *Datenbank Spektrum,* Vol. 17 No. 1, pp. 41-46.
10. **Simpson, M.G.**, 2012. *Making representations: Museums in the post-colonial era*. Routledge.
11. **Sood, A**. (2016). Every piece of art you've ever wanted to see - up, close and searchable. TED lecture. https://www.ted.com/talks/amit_sood_every_piece_of_art_you_ve_ever_wanted_to_see_up_close_and_searchable (accessed 25 November 2018).

Inna Kizhner (inna.kizhner@gmail.com), Siberian Federal University, Russian Federation and Melissa Terras (M.Terras@ed.ac.uk), The University of Edinburgh and Maxim Rumyantsev (m-rumyantsev@yandex.ru), Siberian Federal University, Russian Federation and Valentina Khokhlova (kenzo178@yandex.ru), Siberian Federal University, Russian Federation and Elizaveta Demeshkova (liza-demesh@mail.ru), Siberian Federal University, Russian Federation