Digital cultural colonialism

Citation for published version:
https://doi.org/10.1093/llc/fqaa055

Digital Object Identifier (DOI):
10.1093/llc/fqaa055

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Digital Scholarship in the Humanities

Publisher Rights Statement:
This is a pre-copyedited, author-produced version of an article accepted for publication in Digital Scholarship in the Humanities following peer review. The version of record Inna Kizhner, Melissa Terras, Maxim Rumyantsev, Valentina Khokhlova, Elisaveta Demeshkova, Ivan Rudov, Julia Afanasieva, Digital cultural colonialism: measuring bias in aggregated digitized content held in Google Arts and Culture, Digital Scholarship in the Humanities, 2020;, fqaa055, https://doi.org/10.1093/llc/fqaa055 is available online at: https://academic.oup.com/dsh/advance-article/doi/10.1093/llc/fqaa055/6042086

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.
Digital Cultural Colonialism: measuring bias in aggregated digitized content held in Google Arts and Culture.

Inna Kizhner, Melissa Terras, Maxim Rumyantsev, Valentina Khokhlova, Elisaveta Demeshkova, Ivan Rudov, Julia Afanasieva

Digital Scholarship in the Humanities, Accepted 13th September 2020.

Abstract

In February 2011, Google launched its Google Art Project, now known as Google Arts and Culture, that currently hosts approximately six million high resolution images of artworks from around the world, with an objective to make culture more accessible. We demonstrate that Google Arts and Culture (GA&C) has experienced dramatic growth in recent years and includes artworks for almost every country from the UN member list. However, we document a noticeable lack of balance in the aggregator, with some countries and institutions being prioritized, and a major proportion of the holdings featuring content that resides in the USA. Moreover, after examining Russian and French collections, we see the dominance of artworks from their capital cities in GA&C, while art from provinces is clearly underrepresented. Finally, we find a dominance of art from the twentieth century with some emphasis on artworks demonstrating a break in canon or ‘otherness’ of non-Western countries. The discrepancies that we observe give evidence to support previously posited ideas of digital cultural colonialism. This may be due to a lack of data transparency or availability of previously digitized content in the countries with a lower representation in the portal, or an unintentional digital amplification of conventional traditions of art collection and interpretation that dominate museum displays in larger Western cities. We call for explicit statement by platforms on their collection and selection criteria and the need for researchers to understand the biases emerging from aggregated digitized content.

1 Introduction

Cultural visual data held in memory institutions has become more accessible for enjoyment, education and research over the last two decades: at least from some parts of the world. We have seen a
significant increase in the number of items held in museums and art galleries now available online (Rogers, 2016; Wang, 2016; Hall, 2017). We have also observed improvement in the number of institutions that have introduced open licensing to this data (Kapsalis, 2016; Tallon, 2019; McCarthy 2019). Although these abundant quantities of images are available through a variety of channels, including digital libraries and aggregators of cultural heritage data (Sood, 2016; Petras et al., 2017; Petras and Stiller, 2017), we do not know if and how images in the aggregators represent various cultures, genres and geographic regions (Thylstrup, 2019). This means that if larger quantities of objects, images and stories related to a particular idea or representation of selected knowledge are present in large aggregators of cultural content, these ideas and concepts will be better accessed, disseminated and studied and they will become the foundation of the new digital canon (Earheart, 2012; Feldman, 2017; Thylstrup, 2019; Zaagsma, 2019).

Building a corpus of artworks that include a proportion of objects from a variety of cultures, offering non-colonial social constructs and perspectives, can enlarge the art history canon, and enable further analysis via Digital Humanities techniques (Earhart, 2012; Feldman, 2017, Kizhner et al., 2019). Since the initial museological literature of the 1990s¹ on the reduction of the colonized world presented in museums, an approach that developed in the 19th and early 20th centuries (Boast, 2011, p. 64), numerous studies have emphasized the difficulties of having a multicultural or non-unitary experience in a museum (Ashley, 2005; Knell et al., 2010; Boast, 2011; Polm, 2016). Despite Amit Sood’s (2016) claim that GA&C introduces a new era of accessible art, GA&C has been previously critiqued as a corpus where a number of cultures are underrepresented or marginalized (Pfisterer, 2018). This aggregator is ‘a major producer of cultural content’ with 50 million people visiting its website before 2018 (The Art Newspaper, 2018): it is therefore an important aggregator platform on which to conduct research regarding the structure and organization of visual cultural content.

Literature on pre-existing biases in linguistic corpora (Caliskan et al., 2017), literary datasets (Grieve-Smith, 2019) or collections of museum objects (Huster, 2013) demonstrates how easily bias creeps into datasets and influences the results of further analysis. Literature shows that the bias of the
corpus is strengthened by the pre-digital selection biases of contributing institutions such as galleries, museums, archives and libraries (Huster, 2013), the profit making imperative of publishers (Hauswedell et al., 2020), or biases caused by social circumstances (Korchagin, 2015). Chomsky observed that every corpus is ‘skewed’ (Chomsky, 1959, p. 159) or ‘inevitably biased’ (Clear, 1992, p. 23, see also Raineri and Debras, 2019; Bode 2020). It is noticeable how difficult it is to ‘model away bias’ (Bode, 2020) in literary datasets. Bode (ibid.) suggests that, before moving to quantitative and statistical analysis, some work is needed to produce a curated collection of artworks (or a literary system) that demonstrates and explains its choice and ‘interpretation of documentary evidence’ (ibid., p. 3). Choices that lead to the digitization, publication, aggregation, and promotion of collections, artworks, periods or geographies are often hidden, concealing the limitations of institutional, social and political systems. We argue these epistemic choices should be demonstrated, documented and interpreted.

Bias in the representation of digital collections of cultural content relates to the concept of digital cultural colonialism. Colonialism is sometimes defined as ‘the conquest and control of other people’s land and goods’ under which local communities are ‘unformed or re-formed’ (Loomba, 1998, p. 2). In this paper we understand and define colonialism in its ‘traditional’ sense (as a tension between North and South) and also as an internal phenomenon within countries (as a tension between their centre and periphery). If accumulation of data is indeed ‘a core component of political economy in the 21st century’ (Sadowski, 2019, p. 1), the accumulation of cultural data propels digital cultural colonialism or the ability of online aggregators and other digital publishers of cultural data to represent the epistemologies of dominant communities, reforming the epistemologies of the world outside European or Western culture and ‘appropriating the signs of (a foreign cultural system) to construct their own system’ (Sharp, 2002). Suse Anderson (2020) writes that digital aggregators maintain “the status quo at scale, making claims writ large across millions of objects about whose histories and objects have been worth collecting, and in what circumstances and as part of whose narratives”. We argue that digital cultural colonialism refers to the process of appropriating cultures and communities located outside Western culture, placing them in contexts where they become lesser or produce an
impression of less importance or acquire a veil of ‘otherness’, sometimes self-imposed, in pursuit of constructing and maintaining dominant epistemologies or dominant cultural systems. The concept of ‘otherness’ is understanding of a local culture as dramatically opposed to Western culture, sometimes as a source of inspiration, more often implying ‘binary distinctions between Westerners and racial/cultural Others that support a superiority ideology’, such as ‘advanced/primitive, civilized/uncivilized’ (Caton and Santos, 2008), it sometimes relates to a wish to experience difference in cultures. The concept of ‘otherness’ within the postcolonial theory of Orientalism has been widely covered in postcolonial literature starting from Edward Said’s seminal work (Said, 1978) to the interpretations of Roland Barthes’ complicated image of the Orient as going beyond Said’s simple dichotomies (Sharp, 2002) to more recent papers (Caton and Santos, 2008, Salazar, 2012, Risam, 2018).

Digital cultural colonialism and dominant epistemologies (Santos, 2018; Milan and Treré, 2019) that will result from digital infrastructures of cultural content may lead to statistically biased ‘representations of knowledge and its political subjectivity’ (Knell, 2010, p. 4). This will affect and direct research and user access to cultural heritage, and will, in turn, affect machine learning approaches that build upon existing digital information environments, reinforcing and amplifying biases (Ciecko 2020).

We speak about digital cultural colonialism and not just accessibility issues because GA&C unintentionally produces a complicated and diverse cultural landscape for the countries that are better represented by the aggregator and a simpler, flatter picture for those countries which submitted fewer images or never submitted any images at all. This phenomenon does not result from GA&C’s intentions because participating institutions and (non) participating countries could contribute more images. It is very much related to cultural priorities of institutions and countries, their digitization practices, access and online heritage policies. However it also depends on the editorial principles pursued by GA&C. Documenting discrepancies, imbalances and resulting biases in digitized content is very much the task of an editorial team. In doing so, GA&C can produce content that can become a benchmark for future aggregators of cultural heritage. Otherwise, the aggregator will reflect prejudices and
biases inherited from an imperial era. Unlike exclusion bias (a situation when a place cannot be reached) or selection/collection bias when cultural institutions prefer a particular type of content (Bode, 2018, Bode, 2020), digital cultural colonialism does not only result from the inability of editors to reach certain cultures or places. It is built on collection/selection policies of contributing institutions and depends on exclusion bias, but it is also very much an accumulated and strengthened effect of aggregation. Unlike an edited collection of cultural data with balanced representation and documented decisions, a collection of cultural data edited by algorithms and aggregation leads to digital cultural colonialism due to the necessity to place the data from the Global North, or collected by somebody from the Global North, where the data from the Global South is missing.

To understand the scope and nature of this problem, this paper measures how representative and balanced Google Arts and Culture (GA&C) is, as a major aggregator of digitized content of cultural heritage data from across the world. Since 2011, GA&C has provided a platform which provides the public with high-resolution images of artworks from major galleries and art museums: at the time of writing hosting approximately six million images for the artworks from 1,880 museums worldwide. With an aim to “discover artworks, collections and stories from around the world in a new way. Explore cultural treasures in extraordinary detail and easily share with your friends,” GA&C’s objective is to democratize and share the world’s digital heritage, and its status makes it the ideal locus of study for considering what culture is being supported and made available to wide online audiences. This major initiative, funded by a major technology provider based in the USA, may display biases that have unintentional consequences (and it does not publish any criteria for inclusion of artworks in its portal). We chose GA&C for this study because it is an international platform launched by a highly influential commercial giant that has had a tremendous effect on how information is searched for, sorted, classified and filtered across the world. It has a large collection that at the time of writing covers almost all of about 200 countries from the UN list of member states.

Digital surrogates of artworks, and their aggregators, can be important for a variety of social, political and ethical reasons. They obviously provide access to a great number of major artworks,
including those that are not on display in physical museums, allowing detailed exploration of objects. This has important implications for users who need to remotely access collections (including in a post-COVID19 society). In addition, aggregated digital content is a powerful tool ‘producing the social structures we live in’ (Van Dijk et al., 2018). Online platforms ‘come with specific norms and values inscribed in their architectures’ (ibid., p. 3). Furthermore, literature related to online platforms places importance on the private interests of data collection, aggregation, ownership and profit-making linked to the political economy of data (see, for example, Van Dijk et al., 2018, Sadowski, 2019, Sadowski, 2020, Lane, 2020). Mansfield (2014) shows that GA&C, although a non-profit, non-commercial division, is still a part of a huge commercial giant with all the possible implications of data colonialism (Couldry and Mejias, 2019). Another complication is that online aggregators use the data received from institutions encumbered by the social constructs of empires and historical colonialism (Karp and Lavine, 1991; Knell et al., 2010; Boast, 2011) and strengthened by big data structures (Fylverbom and Murray, 2018).

Cultural colonialism of online aggregators and the bias introduced as a result of aggregation or inherited from participating institutions can be measured by calculating how each country is represented. We can measure the number of objects for each country (these objects may or may not be in the country’s institutional collections, for example, they may be Egyptian objects in the British Museum) and the number of objects and collections representing a country, by calculating institutional involvement in publishing digitized cultural content through GA&C. This may be able to identify biases caused by uneven digitization policies or open access and online heritage policies, or political and social constraints.

In this paper we demonstrate the global level bias when a few dominant countries are represented by the largest amount of images, showing Global North vs Global South tension. We reveal a similar tension using as example of two chosen countries, Russia and France, where the center - periphery tension repeats the tension of the first layer. The third (micro)layer is the tension of collec-
tion bias. In the case of Russia and France, this third layer shows that art sent to the aggregator introduces colonial themes. In some other cases the third layer substitutes the cultural signs of the Global North for the representations of a country, as in the case of Kazakhstan and North-American astronauts (see below).

We start by adapting the concept of representative and balanced digital collections from corpus linguistics to aggregated visual content. We then examine how Google Arts and Culture organizes its digital collection. Can we really observe discrimination and Western bias in the representation of countries and contributing institutions? If yes, what cultures or countries are represented to a greater extent? Does it extend to ‘foreign’ exhibits in national collections, for example, Egyptian objects in European museums? We reveal discrepancies in the number of objects from the top five countries with the largest proportion of displayed items (USA, UK, Netherlands, Italy and South Korea) and the rest of the world. Moreover, countries represented by GA&C by the items kept in other countries’ institutions do not demonstrate balanced representation. The first nine countries in the ranking account for about 40% of the GA&C collection.

We then look at the geographical representation, types of art and time periods for the items published in GA&C for two large countries: Russia and France, chosen given the availability of national level datasets regarding their collections. We compare their representation in GA&C to the data presented by national museum aggregators in the two countries. We observe that the data are skewed towards art objects from capital cities in both countries, with art from provinces represented to a lesser degree. The art presented by GA&C is related to contemporary art and photography from the 19th-20th centuries that is meant to show changes in visual canon and demonstrate the ‘otherness’ of non-Western cultures. We identify this by comparing the dominant types of art and time periods with those from national aggregators and by looking closely at the items representing five top artists for the collection of each country (Russia and France). We show that lower representation of institutional collections for some countries in GA&C (as compared to the top five countries) is not connected with digitization policies, at least for the countries examined, as large quantities of objects from Russian
and French museums have been digitized (probably accompanied with metadata of low quality). Lower representation in these cases can be related to policies of open access and lack of emphasis on online dissemination of museum images: in the case of Russia, museum data dissemination and reuse for non-inventory purposes are not a priority (Kizhner et al., 2019). We cannot generalize these results to all countries from the UN list as we do not know the rates and scope of digitization in these countries and whether low levels of digitization can be an obstacle to sending images to international aggregators. If we consider the countries that have never included images in GA&C, we can see that Kazakhstan, a country in Central Asia, at the time of writing, is mainly represented through NASA photos. These photos depict American and Russian astronauts preparing for an American-Russian (Soviet) space mission which started in Baikonur, where the Soviet space station was located in Kazakhstan. This is a result of NASA uploading a large amount of photos and Kazakhstan’s cultural institutions providing none. We can see how a country (Kazakhstan) is represented through a part of a North-American cultural system as an effect of aggregation. This is important because it shows how unintentional aggregation effects lead to the colonial center-periphery appropriation of representations, where a geographical place represents Western achievements.

Finally, we link these results and supporting evidence for discrimination and digital colonialism in aggregated digital content with the imperial bias inherited from the print era that is obviously opposed to the views implicitly expressed by GA&C. Amit Sood, director of the project, makes an emphasis on diversity in his TED talk (Sood, 2016), saying that along with European paintings they have the ‘Black history’ channel and samples of Japanese craftsmanship. Sood also speaks about temporal diversity of their collections and about the ability to ‘spend hours looking at every decade and the contributions in that decade’ (ibid.). The visuals shown in this talk imply diversity of countries and experiences. The diversity in the visual content presented to viewers on the front pages of GA&C’s platform and in Sood’s talk is in contrast with the selection bias inherited from memory institutions that may be further amplified by data transparency, availability of previously digitized content that can be contributed to GA&C and open access or licensing policies for some countries.
and contributing institutions, making it easier for digital platforms to work with some institutions from some parts of the world. These effects can be strengthened by search engine algorithms and this has further implications of the necessity for clear understanding of biases emerging from aggregated digitized content and the need for explicit documentation by platform providers on their selection and featuring policies.

2 Data Bias in Related literature

Van Dijk et al., (2018) question the value of online platforms for public systems and ask ‘who…should be responsible and accountable for anchoring public values in the platform society’. While the value of online platforms aggregating cultural heritage for crowdsourcing projects such as Wikimedia Commons or Flickr Commons has been widely discussed (see, for example, Kalfatovic et al., 2008; Oomen and Aroyo, 2011; Ridge, 2014; Terras, 2016), scant literature examines social implications of national and international aggregators of visual cultural heritage. Stiller and Petras (2018) note that forty evaluation studies for Europeana Collections include only two papers on social constructs that determine Europeana’s social value. In particular, Valtyssoon (2012) shows how Europeana continues the tradition of memory institutions’ authoritative voice. Yankova et al. (2015) outline social requirements related to Europeana content such as more balanced content with the minimum records from each EU Member State.

In this paper, we use the concepts of representativeness and balance as they are used in corpus linguistics (see, for example, Biber, 1993; Leech, 2007) where a corpus is representative when it contains ‘all parts of the linguistic variety’ (Gries, 2009, p. 1231). A balanced corpus should follow the rule ‘that the proportion with which a particular part is represented in a corpus should reflect the proportion the part makes up in this variety and/or the importance of the part in this variety’ (ibid., p. 1231). Gries (2009) explains that it is difficult to understand the proportion of the part in a real life
phenomenon, such as a language or, in our case, collections of artworks where the size of the population or its stratification is unknown. It is no less difficult to understand the importance of these parts (Leech, 2007; Korchagin, 2015).

Although every corpus is ‘inevitably biased’ (Clear, 1992, p. 23) and ‘the Holy Grail’ of a representative and balanced collection of electronic texts (or images) is unlikely to be reached (Leech, 2007), linguists that debated the concepts of representativeness and balance at the end of the 20th century insisted on a detailed discussion of methodological principles: how materials included in a corpus are obtained and whether sampling methods can influence the validity of results (Woods et al., 1986, cited from Clear, 1992). However, a recent study (Egbert, 2019) shows that the majority of linguists who published empirical corpora studies in influential linguistic journals in 2014 believed that the most important characteristic defining a corpus was its size and not its representativeness. This has important implications not only for linguistic studies based on empirical corpus research but also for a variety of studies that use large datasets, including those with digitized cultural content.

Our hypothesis is that, under the conditions of aggregated digitized content, neocolonial dominance (Boast, 2011) is revealed in representing arts and culture from the countries that submitted, or allowed or selected to display more digitized images within an aggregated content. Although much has been written about social constructs and colonial attitudes for physical collections of artworks, little is known what happens when the selection happens at the level of aggregators or online platforms such as Google Arts and Culture and the ‘fabrique’ of selected objects and stories (Knell, 2010, p. 7) produced by museums is augmented through mediated selection. We aim to investigate this here.

3 Data

3.1 Data Source
Our primary source of data was from the GA&C platform itself, which publishes sets of high-resolution images of artworks, videos and panoramas delimited by their particular place of manufacture, such as the United States of America, China, France, Russian Federation, Paris or London. Fig. 1 demonstrates the rates at which a variety of cultural institutions joined GA&C since its establishment in 2011 by the Google Cultural Institute. We can observe a steady growth between 2009 and 2015 and a sharp increase in the number of contributing institutions starting after 2015. The rise can be explained by the wide popularity of platforms at that time, the growing popularity of GA&C due to the way it is known to drive audiences towards collections (BrodeFrank 2020), and the fact that some institutions started sharing their collections after adopting open licensing policies, or carrying out their own digitization programs that gave them the assets to contribute to GA&C. Our analysis is based on the full collection of two-dimensional images with metadata available on GA&C web site (ca. 5,500,000 records) as of June 15th, 2019. All our numeric estimates below are based on Collections (museum collections or collections of other holders, such as LIFE Magazine or Opéra national de Paris, as defined by GA&C) for each country from the list of the United Nations’ member states. For the data on foreign museum objects in other countries’ institutions, we used the ‘Discover this place’ section that lists all the items mentioning a country (or a city in this country) in metadata or free texts accompanying images. The ‘Places’ category chosen from the menu gives an opportunity to choose a particular place (e.g. France) and study how the country’s artworks are represented in GA&C. Studying the ‘Places’ category allowed us to see how collections of particular countries are represented through GA&C.
We identified the number and size of the collections submitted to GA&C and the country of collection holders. Looking at the page of a country we can see the number of collections (e.g. 87 collections for France, as of October 30th 2018). Each collection page displays the number of items in the collection\(^9\). By adding up the number of collections in a country and the number of artworks from each collection we get the numeric estimates for analysis and comparison. As of October 30th, 2018, our GA&C data set contained ca. 5,500,000 images for the artworks from 1,780 collections across the world.

The number of collections and artworks per se is not exactly meaningful if we do not know the number of artworks in a given physical collection or the number of artworks in a country or, indeed the number of collections in a country. However, the number of digitized museum objects for some countries can be checked because it is included in national statistical and governmental databases. To
that end, we chose two countries for sampling and further analysis: Russia and France, selected because both countries have an impressive number of museums and museum objects - ca. 2,300 museums, 80 million museum objects for Russia (Kizhner et al., 2019) and 1,224 institutions, 121 million museum objects for France\(^{10}\). Museums from both countries are listed in the study of 18 most influential museums across the world (Van Riel and Heijndijk, 2017). Furthermore, statistics regarding the holdings of their museums are available from government sources: the State Catalogue of the Museum Collections of the Russian Federation (RF)\(^{11}\) and the Open Platform of Heritage published by the French Ministry of Culture\(^{12}\) which would allow comparison with how these collections are represented on GA&C. The Ministries of Culture of both countries maintain the databases of digitized artworks from national, regional and municipal museums. The databases are published online and records can be filtered to explore the number of museum objects of different genres and time periods. This enables us to find out how representative GA&C is in terms of geographical distribution or distribution across time and genres. We used a qualitative methodology to study the principles of GA&C’s content selection. We looked at the images’ accompanying texts published on GA&C platform written by the curators of French and Russian institutions from the lists of these countries’ Ministries of Culture\(^{13}\). In addition, we used standard information seeking procedures of starting, chaining, browsing, differentiating, monitoring, and extracting (Ellis, 1989) to find evidence of what determines content selection for GA&C participating institutions.

The top national collection in GA&C comes from institutions located in the USA and it significantly exceeds all the other national collections (Fig. 2). However, GA&C presents French and Russian collections generously enough to get an impression of their richness and variety. They are not the countries with the highest number of collections in GA&C, nor are they the countries with the lowest number of collections and artworks (Fig. 3, Fig. 4).
Fig. 2: The top 5 countries presenting collections of their cultural institutions in GA&C (by item, by the country in which the institution who hosts the item is located).

Fig. 3: Russia and France are among the top 20 countries presenting their cultural heritage in GA&C by location of cultural institutions. The graph shows the bias of the USA’s collections in GA&C online aggregator.
Fig. 4: Russia and France are among the top 20 countries presenting their cultural heritage in GA&C by number of artworks by location of cultural heritage institution that holds them.

3.2 Identifying Russian Collections in GA&C

We calculated the number of collections and artworks from all Russian cultural institutions published by GA&C. Our next step was to find out how many Russian museums had digitized (parts of) their collections. GA&C includes samples of artworks from public museums. It also has collections of private museums, collections of individual artists and collections of other cultural institutions such as theatres, research institutes, archives of photo magazines, etc. The number of these institutions may be equal to or may even outnumber public museums reporting to public authorities. The number of artworks in GA&C’s Russian collections from public museums was 2,844, while the number of artworks from other cultural institutions was 2,784. We restricted further analysis to the artworks and collections of public museums. By doing so, we can compare the geographical distribution of Russian
collections in GA&C with that in the database of digitized collections of public museums provided by the State Catalogue of the RF Museum Collections. The rates of digitization in Russia in autumn 2018 are known because the State Catalogue of the RF Museum Collections published by the RF Ministry of Culture\textsuperscript{14} shows the data on the number of available objects with images.

In spring 2019, we used the dataset of Russian artworks from public museums to analyze how artworks are represented by GA&C digital platform. We identified and mapped the geographical location for each museum to compare the dataset presented in GA&C with the list of museums that have already digitized (parts of) their collections (Fig. 5b). The list of museums with digital images in their databases was found on the web site of the State Catalogue of the RF Museum Collections\textsuperscript{15}. We also tabulated each record from the Russian collections from GA&C in a spreadsheet that included the type of an object, time period of its creation and its author. The classification schema for object types, such as paintings, sculptures, documents or coins, is present in the State Catalogue and it is one of the two facets for browsing its web site. This enabled us to compare the numbers for object types in our data and the numbers in the groups of collection types from the State Catalogue to find out whether they are proportionate. We identified collection types for the artworks published by GA&C using metadata and free text accompanying images. Sometimes we had to use our own judgment as GA&C does not maintain a metadata standard (Mansfield, 2014) and ‘collection type’ field or a free text were not always available. Our next step was to identify the creators or authors whose works were dominant in GA&C’s Russian collections. This was done by measuring the number of times a name appeared in our data for 10 most frequent names. It was impossible to compare this with the data from the State Catalogue as this publishing platform does not provide a list of most frequent names of creators or authors. Overall, during the first stage of the analysis when we looked at the number of objects in the collection, we studied 49 Russian collections with 5,628 images. After excluding private museums and non-museum institutions from the analysis, we worked with 32 Russian collections and tabulated data for 2,844 objects.
3.3. Identifying French Collections in GA&C

We calculated the number of collections and artworks from all French cultural institutions published by GA&C in autumn 2018. We identified the location of these collections and tabulated each artwork. The Excel spreadsheet included such data as the name of the author, time period, collection type, and availability of metadata. To compare the geographical location of museums, we obtained a list of French national museums from the website of the French Ministry of Culture. Comparing the data across authors and time periods was possible because the French Ministry of Culture publishes the Joconde database that includes images for ca. 400,000 digitized museum objects from French museums reporting to the French Ministry of Culture. Object types presented by the Joconde database...
were too numerous to group within a reasonable number of sets so this comparison was excluded from the analysis. Overall, during the second stage of the analysis in spring 2019, we worked with 94 collections and 19,024 museum objects, with their digital copies published through GA&C platform. After comparing the list of French cultural institutions in GA&C and the list of French national museums, we excluded museums and collections of non-museum institutions that were not in the list from the second stage of the analysis (geographical distribution of national museums represented in GA&C, compared across time periods and authors). This left us with 21 collection and 6,575 museum objects for further analysis and comparison.

We used the method of frequency-based collages (Junginger et al., 2019) to visualize our results for Russian and French collections (top creators). The maps showing the geographical distribution for the GA&C collections in general and Russian/French collections, in particular, were produced using Adobe Illustrator.

4. Results

4.1 Is Google Arts and Culture Representative and Balanced?

Fig. 2 shows the top five countries with cultural institutions featured in GA&C. We can see that museum objects from the USA, UK, Netherlands, Italy and South Korea account for 93.4% of the content published by GA&C platform, with the USA alone accounting for 82% of images found on the site. The share of the other 190 countries from the UN list is 6.6%19. It means that the probability of images coming from these countries for neutral search terms such as ‘chess’, ‘glove’ or ‘earring’ is much higher than for the images coming from the rest of the world, prioritizing these cultures or rather their museum interpretation practices (especially those of the USA) above others.

Rather than studying the evenness in representation at the level of countries, we approach it at the level of cultural institutions or collection holders that publish images of their objects via GA&C
platform. Fig. 6 shows top five cultural institutions with their collections in GA&C. We can observe that LIFE Photo Collection\textsuperscript{20} is fifteen times bigger than the next collection in the ranking and accounts for 76\% of cultural objects published via GA&C platform (and also shows how the inclusion of one large collection can skew representativeness across the board). This means that a share of images from this collection for every query related to the twentieth century such as ‘Margaret Thatcher’, ‘nazi’ or ‘collective farm’ may be higher than a share of images from other cultural institutions, and also be presented via a North-American cultural lens.

![Fig. 6: Top 5 cultural institutions presenting their collections in GA&C (as of June 15th, 2019)](image)

Although our results show huge differences in the number of objects for different countries at the level of geographical distribution of cultural institutions, we cannot say that GA&C is not a representative corpus as foreign objects in other countries’ institutions can be retrieved almost for every country. One hundred and twenty three countries without the institutional representation in GA&C are represented through their countries’ objects in other countries’ cultural institutions (see Appendix 1). This happens almost for every country from the UN list, starting with Vietnam with 0 objects in the \textit{Collections} section and almost 20,000 objects in the \textit{Discover this place} section, to Liechtenstein and Tajikistan which both have zero objects in the \textit{Collections} section and 13 objects in the \textit{Discover this place} section (see Appendix 1). However, our results show that, even for that group, the first country in the ranking (USA) outnumbers the other countries so much that one of the four images a
user expects to see in the aggregator may include some metadata related to this country (Fig. 7), unless these results are manipulated through GA&C’s algorithms.

Fig. 7: Geographical distribution of digitized content in GA&C (metadata mentions a geographical name as a place of the event in an image or as a place where the creator was born, etc.)

Even if we exclude the LIFE Photo Collection from the analysis related to the representation of cultural institutions in GA&C (Fig. 8), we get similar results, with the USA collections being at the top of the ranking, especially if we consider that new collections are being added to the platform. Our data show that only 72 countries out of the 195 countries from the UN list publish samples of their institutional collections via GA&C (see Appendix 1). This is not the case for 123 countries, which means that GA&C is not a representative or balanced aggregator in terms of representing cultural institutions. When we look at foreign objects in other countries’ institutions (Fig. 7) we see that GA&C content is much more representative and, probably more balanced. We can see that almost all the countries from the UN list are represented through metadata or free text or a city where an artist was born. However, the number of objects related to the top nine countries, ca. two million objects
(Fig. 7), shows that we cannot say that GA&C platform is a balanced corpus of visual content as the nine countries account for about 40% of the platform’s visual content.

![Geographical distribution of cultural institutions with digitized content in GA&C, excluding content found in the LIFE Photo Collection.](image)

When we look at the web sites of the top five ranking institutions we can see that three of them - the Natural History Museum in London\textsuperscript{23}, the Metropolitan Museum in New York\textsuperscript{24} and Rijksmuseum in Amsterdam\textsuperscript{25} - offer to access and download their data using the API\textsuperscript{26}. This means that previously digitized assets can be easily reused, and that museums can build new partnerships with technology companies, including GA&C, in particular to study how to use visual data for research, industrial and commercial applications (Tallon, 2018; Valeonti et al., 2018) and for brand licensing to use images in product lines (Kapsalis, 2016). Perversely, the drive for open licensing and OpenGlam\textsuperscript{27} in certain countries (predominantly the USA and Europe, see Terras, 2015; McCarthy and Wallace, 2018) may further be prioritizing the importance of their collections in the new digital canon.
4.2 Are Russian Collections in GA&C Representative of the Country’s Geography or Art Genres?

Our results so far show that GA&C is not balanced at the level of the world geography. Next, we would like to learn more about the geographical distribution of cultural institutions within two single countries, Russia and France. As of June the 15th, 2019, collections of Russian national, regional and municipal museums in GA&C were represented by 32 cultural institutions and 2,844 objects. When we compare their geographical location with the locations of 2,189 museums registered with the State Catalogue of the RF Museum Collections (Fig. 5), we can see that GA&C publishing activity was restricted to museums in the central part of the country and its capital city Moscow. Fig 5a shows that GA&C represents seven Russian provinces or 6% of eighty five regions in the country. Fig. 5b shows that the geographical distribution of Russian museums with digitized content is much wider and it covers almost all the regions in the Russian Federation.

To find out if the Russian collection of the aggregator is representative in terms of time periods, we looked at the year when an artwork published on GA&C platform was created. As we mentioned in Section 3, the State Catalogue does not have a filter that could help us to compare distributions across time periods. Our results only show the distribution of time periods for the objects from Russian collections published via GA&C platform (Fig. 9). As we can see, the twentieth century artworks account for almost 60% of the collection, while the objects created before the twelfth century are represented by less than 1% of all images. As shown in Figure 10, paintings, photos and decorative art account for half of the images from Russian museums in GA&C, while weapons, natural history and sculpture have low representation. Paintings are clearly overrepresented in GA&C as compared with the State Catalogue (Fig. 11). Our results for the top five creators in GA&C show that they are twentieth century artists, fashion designers, poets and photographers. Figure 12d displays a newspaper cutting that describes how Leon Bakst, costume designer for ‘Ballets Russes’ in the early twentieth century, influenced the development of fashion in Europe and North America. The newspaper
article belongs to a series of cuttings submitted to GA&C as a part of content solicited from museums to celebrate the Year of Fashion in 2016 (Shumanova et al., 2019). The cuttings account for a major part (40%) of the museum collection. The museum curators write: ‘Exclusively for the Year of Fashion project by Google Art Project, we have looked through Bakst’s archive to find clippings dedicated to fashion’ (Shumanova et al., 2019).

Fig. 9: Temporal distribution of Russian digitized content in GA&C by date of creation

Fig. 9: Temporal distribution of Russian digitized content in GA&C by date of creation
Fig. 10: Types of art from Russian collections in GA&C by media

![Bar chart showing the percentage distribution of art media in Russian collections in GA&C.](chart.png)

Fig. 11: The number of paintings in GA&C is much higher than in the State Catalogue. Types of art from Russian collections in the State Catalogue and in GA&C

(a)  
(b)  
(c)  
(d)  
(e)
Our results show that the distribution of artworks across time periods is uneven and prioritizes the twentieth century. The State Catalogue of the Museum Collections (see Section 3.2) does not show the distribution of items in its databases by the date of creation so we cannot say whether GA&C Russian collections deviate from the proportions represented by the national portal. However, our results suggest that users are more likely to see twentieth century paintings and photos in the Russian part of GA&C collections than artworks of other genres or time periods.

4.3 Are French Collections in GA&C Representative of the Country’s Geography or Time Periods?

As we demonstrated in Section 3, France is an appropriate country to prove that the results observed for Russia are not country-dependent but rather the editorial principles that govern how the GA&C platform presents its content. The results follow from the intentional choice of participating institutions (including that of the museums that decided not to go into the partnership with GA&C) and the editorial principles that govern how the GA&C platform presents its content. When we exclude non-registered museums and non-museum institutions from the analysis, we are left with 21
French museum and 6,575 museum objects published via GA&C platform. The differences in geographical distribution of museums represented in GA&C and museums from the list of the French Ministry of Culture are presented in Fig. 13. We can observe the disproportionate prevalence of museums from Paris in the GA&C French collection and complete absence of museums from a number of provinces, especially those located further from Paris. As one can see, Île-de-France, one of the most important and centrally located French regions including Paris, is represented by eleven museums or 52.4% of museums in the sample. Paris is represented by eight museums or 38.1% of all images. France as a country is underrepresented as the GA&C collection only includes museums from eight provinces or 44.5% of the eighteen provinces in the country. All in all, Fig. 13 displays the trend we have already observed with Russian collections in GA&C: museums represented are located in the capital city or in the region that includes the capital city, and provinces located at a distance from the capital are underrepresented or not represented.

Fig. 13: Geographical distribution of French digital collections in (a) GA&C, excluding private museums and non-museum institutions (21 museums and 6,000 images); (b) Ministère de la Culture (France) (1,220 museums)
Fig. 14 shows the comparison of the distribution of artworks across time periods in GA&C’s French collections and in Joconde database (see Section 3). Blue columns for GA&C clearly show the dominance of the twentieth century (41.7%) followed by the nineteenth century (36.5%).

Fig. 14: Temporal distribution of digitized content from French museums in GA&C

Fig. 15 shows the samples of works of top five creators featured in collections published by GA&C. The creators are twentieth century fashion designer and nineteenth century photographers. The French fashion designer Margueritte Porracchia’s artworks account for 1,200 images or almost 20% of French collections published by GA&C platform. They date back to the first part of the twentieth century and their display by GA&C may relate to the Year of Fashion, while the museum curators may have submitted works that demonstrated a change in fashion canon related to cultural transfers in the early twentieth century, in particular related to orientalism as the ‘invention of other geographical spaces’ (Said, 2000). The second most frequent creator is Pierre Tremaux (Fig. 15a), a nineteenth century French photographer collecting visual documentary evidence of African and Asian
continents who, as Addleman-Frankel (2018) suggests, was a part of the colonial culture phenomenon.

Fig. 15: Top five creators in French digital collections (GA&C) are 20th century fashion designer (1,300 images or 20% museum objects in the sample) and 19th century photographers; (a) Pierre Trémaux, Voyage au Soudan Oriental, 1847 - 1853 © musée Nicéphore Niépce, Ville de Chalon sur Saône; (b) Alfred Nicolas Normand, CHARITE-s-Loire. La rue des Hôtelleries, 1892 © musée Nicéphore Niépce, Ville de Chalon sur Saône; (c) Marguerite Porracchia, dress and coat, sketch, 1947 © MAD, Paris, musée des Arts décoratifs; (d) Benoît de Tyskiewicz, Mlle Marie Chave, 1890 © musée Nicéphore Niépce, Ville de Chalon sur Saône; (e) Chambre équipée d'un diaphragme à iris, utilisée par Joseph Nicéphore Niépce, 1816-1833 © musée Nicéphore Niépce, Ville de Chalon sur Saône. Images accessed 10.02.2020, https://artsandculture.google.com/entity/france/m0f8l9c?categoryId=place

4.4 Principles of Content Selection in GA&C
The question of whether Google Cultural Institute recommends what type of content is desirable for aggregation can be answered very broadly. Generally speaking, the decisions made by both Google and their cultural institutions regarding what to feature may be chosen with the belief that their users are interested in popular content and that cultural institutions submitting images ‘are catering to a different audience’ (The Art Newspaper, 2018). Sometimes they may solicit content that suits a particular project or theme such as ‘Year of Fashion’ (Section 4.2 above). We cannot generalize from here that every museum is instructed what kind of content to submit but the process of content selection seems to be in line with Elizabeth C. Mansfield’s observation (2014) that GA&C is a producer of non-scholarly ‘infrastructures of knowledge-making’ (Mak, 2017, cited from Bode, 2020). The selection practices are not openly published or widely known, and it may indeed be the case that institutions are contributing the type of generalized content they believe should be present on GA&C, for a general audience: thus unintentionally compounding biases present in the platform.

5 Discussion

We can distinguish two main issues arising from our results, and their limitations. First, do they imply new digital cultural colonialism when aggregators prioritize certain cultures and periods at the expense of other cultures? Second, can they inform the principles that should be used to organize large datasets of digitized cultural content?

5.1 Limitations

Although our results show that GA&C prioritizes its top countries in terms of the number of images at the expense of the rest of the world (Fig. 8, Appendix 2), we do not know if Google attempts to be more representative when results for real queries are displayed or how much is dependent on the type of search, or location, of search. Metadata are not consistent over the GA&C dataset, but
simple queries may match results from a variety of cultures and museums. Do political terms and
terms related to recent history, such as ‘Margaret Thatcher’, ‘holocaust’ or ‘collective farms’ return
more images from a narrow set of countries and institutions as compared to broader terms such as
‘horse’, ‘vessel’ or ‘earring’? How diverse will the results for broader terms be? Are they going to
cover top twenty countries in GA&C or are they going to include the rest of the world? What are the
rates of inclusion for each country, and does the IP address location affect the results displayed? How
are the institutions, and Google itself, using analytic features of the platform to further drive search,
or institutional strategies on display and audience engagement (Filippini Fantoni et al 2012, Stein
2014)? Our paper provides results for the number of images submitted to GA&C by cultural institu-
tions across the world. It does not report on what happens with the data when they are displayed as
search results and whether this makes the aggregator more representative and balanced: the search
algorithm is as important as the content, although remains a blackbox (with the search algorithms
themselves demonstrating known problematic biases in prior research (Noble, 2018)).

Our results on the distribution across time, art genres and dominant artists only refer to the
images of artworks within collections in France and Russia. These are not sufficient to generalize
about the content of the aggregator as a whole. On the other hand, the LIFE Photo Collection (see
Section 4.1) accounts for 76% of GA&C’s content (Fig. 8), which may make the aggregator biased
towards presenting 20th century photography for some search results.

We do not know the distribution of requests that GA&C sent to museums across the world to
submit images to the aggregator, nor its processes for deciding on inclusion. Neither do we know the
geographical distribution of non-response to compare it to our results on the geographical distribution
of Google cultural collections. We also do not know of any forward-thinking plans which may be in
place to address any issues of geographical bias. While we hear from our institutional contacts that
Google requires institutional resource to upload assets to GA&C, and that choices are to be made by
institutions, we do not know if any choices are encouraged or vetoed, and issues of agency and unin-
tentionality with regards to institutions are unclear. We compared our results for French and Russian
collections with the national aggregators of cultural heritage images. While the State Catalogue of the RF Museum Collections with its 20% of the country’s museum objects from all Russian regions can be considered a representative sample, the French database of museum objects includes ca. 400,000 images or 0.3% of museums objects in French collections that store over 121 million museum objects (Joconde, 2019). Such a sample can be just an approximation of the situation observed for French museum collections. However, this does not change our conclusion that GA&C displays more twentieth century artworks than objects from any other periods (Fig. 9, Fig.14).

5.2 Digital Cultural Colonialism and Aggregators of Digitized Content

Our results support existing postcolonial and museological literature, moving their conclusions on cultural neocolonialism to the digital realm and showing how they are strengthened through aggregation. This is supported by the global level discussion and it is also supported by the second level where we study France and Russia. It is especially important for Russia, as the colonial status of some of its provinces underrepresented through GA&C such as Siberia or the Far East, have been previously noted (Forsyth, 1992, Suslov, 2007, Suslov, 2017, Popodko and Zimnyakova, 2018). We show that distant provinces in both countries are unintentionally excluded from representation. This may happen because distant provinces are not motivated to be a part of global representation. This may be also due to colonial practices when a significant part of the local income is appropriated by the centre with a consequent lack of funding to support social and cultural development (Popodko and Zimnyakova, 2018). It also shows how easily bias, be it exclusion bias or colonial bias, or the combination of the two, finds itself in online aggregators. The third level of our analysis shows collection bias privileging modern art with colonial themes in the institutional collections of Russia and France submitted to GA&C. Literature on neocolonialism in presenting multicultural contacts in museums (Boast, 2011) also explains our results on the existence of autoethnographies and colonial images in French and Russian collections. Specifically, we discuss how existing literature is relevant to our
results in terms of 1) geographical distribution and 2) the distribution of dominant authors in French and Russian collections.

The results demonstrate biases in the distribution in the number of objects of the countries submitting images to GA&C (Fig. 8) and biases in the distribution of the institutions that hold cultural objects (Fig. 6). This is in line with the literature that shows how digital colonialism and resulting data structures privilege ‘the expectations of a dominant technological culture’ (Loukissas, 2019, p.10; Risam, 2019). The dominance of the objects from American, British and Dutch cultural institutions may be a consequence of data transparency (Hansen and Flyverbom, 2014; Schudson, 2015; Albu and Flyverbom, 2016) which as Michael Schudson argues was introduced in the USA as the openness of government information in the 1960s (Schudson, 2015) and a result of ‘the broad distrust of convention and authority’ (ibid., p. 27). Another reason for the abundance of data from Western cultural institutions is that for some of their previously digitized content data can be accessed for reuse using APIs (see Section 4.1), which makes digital assets easily available for the general public and companies that become their partners.

The ubiquity of images from the top institutions in GA&C’s ranking (Fig. 6) creates the critical mass of concepts, attitudes and ideologies which, by means of their quantity, push out other cultures and ideologies from interaction. This means that the postmodernist ‘plurality’ of voices (Connor, 1991, p. 80 cited from Harrison, 1997) which could have been GA&C’s intention (Sood, 2016) was enacted in a way that resulted in augmenting colonialism and strengthening prejudices and canons (Earheart, 2012, Zaagsma, 2019) embedded in the colonial institutions of the past. Postcolonial and postmodernist ‘assumption of change’ (Harrison, 1997, p. 41) was supposed to lead to a multicultural and multivocal collection of cultures, regions and periods in a digital world, greatly enlarging the information outreach via GA&C’s platform. This did not result in the ‘enactment of change’ (ibid.), a result which was likewise not observed in a change in the physical museum (Harrison, 1997; Knell, 2010; Knell, 2018).
This kind of discrimination is strengthened through the dominance of artworks from central museums in Russian and French collections and, presumably, throughout the aggregator. For Russia, the centre-periphery dichotomy, with most objects representing museums from central regions is in agreement with literature on the colonial nature of economic relations for some of the country’s provinces (Suslov, 2008, Popodko and Zimnyakova, 2018) and imbalances in the spatial distribution of economic growth (Suslov, 2018). In addition, the dominance of 20th century images in the aggregator (Fig. 6) results in ‘imagined communities’ (Anderson, 2006) constructed through the ‘invention and construction of a geographical space’ (Said, 2000, p. 181). An example is Leo Bakst’s archive of newspaper cuttings (Fig. 12d, Section 4.2) that shows how the artist and stage designer for ‘Ballets Russes’28 early in the 20th century influenced the Western fashion through his work with fashion designers from London and Paris. Marten-Finnis (2013) shows how Bakst shaped “the Western perceptions of the Oriental Other” (ibid., p. 276) contributing to Western expectations of how the Oriental should be represented. The curators who submitted Bakst’s archive to GA&C returned to Bakst’s heritage in the context of sending solicited digitized content as a part of the Year of Fashion project, building a new layer of perceptions associated with the Orientalism (Said, 1978). This becomes a remembrance of the colonial environment of the early 20th century in the new context of GA&C’s Year of Fashion. However, a simple explanation for the dominance of 20th century artworks can be that it is a response of submitting institutions to predicted user expectations and a higher demand for contemporary art from modern consumers. In addition, curators can be driven by a wish to match user expectations and user behavior observed through web analytics to measure audience engagement that is routinely used across the museum sector (Peacock, 2007, Chan, 2008, Filippini Fantoni et al 2012, Stein 2014, Yi et al., 2020).

The fact that the aggregator privileges contemporary art (Fig. 9, Fig. 14) can mean that major parts of museum collections in Russia and France hold art and photography from the early 20th century up to the present time. This may be representative in terms of production because contemporary objects and 20th century photography may be abundant in museums. With the dominance of 20th
century art, GA&C, however, is unlikely to present a balanced collection of images in terms of cultural importance or cultural distinction both within the canon and beyond it. We do not know the distribution across time periods for digitized Russian artworks and we do not know the proportion of contemporary art for other collections of digitized artworks across the world, although a number of major national museums provide this functionality through their web sites’ search engines. If Joconde, the French national aggregator, reflects the importance of collections or cultural distinction or perception of art (Fig. 14) in its preference for the 19th century art this can be its editors’ choice and it is not documented on the Joconde web site. The dominance of contemporary art in GA&C’s French and Russian collections can be down to curators’ choice and marks a move from expert classifications across time periods in art history (Knell, 2018) to focusing on the changes in the canon of art history that happened between the late 19th century and the present time. This is further evidenced through our results on the most frequent creators for Russian and French collections in GA&C (Fig. 12, Fig. 15). Changes of canon at the turn of the twentieth century and early decades of the twentieth century resulted in cross-cultural transfers and included, among others, photography as an art form (Beaulieu and Roberts, 2002), photography as a reproduction medium promoting colonialism and/or cultural transfers (Osborne, 2000; Ryan, 1997; Behdad and Gartlan, 2013), changes in fashion connected to new European demand for inspiration (Martin and Koda, 1994) and changes in Russian literary style that was evidenced through the change in poetic rhythm and meter (Ehrlich, 1952; Gasparov, 1995). All these are represented by the top five creators for French and Russian collections. A considerable part of the most frequent creators from GA&C’s Russian and French collections represents colonial and oriental ideas (Marten-Finnis, 2013, Addleman-Frankel, 2018, Imbert, 2020). Addleman-Frankel (2018, p. 50) shows how 19th century colonial photography, in particular, Pierre Tremaux’s photos (second top creator in French collections), exemplified a move towards objectivity and competed with other graphic media as illustrations in colonial geography books. This was, undoubtedly, an interesting turn in the canon that curators of holding institutions may have found suitable for an international aggregator.
The dominance of images from capital cities and provinces surrounding capital cities in Russian and French collections also contribute to our understanding of the aggregator as a corpus that inherits imperial bias of the printed era. On the other hand, we do not know if museums from provincial regions were ever approached by GA&C and whether their absence in the aggregator relates to their intention to control collections from appearing on external platforms and keep them closed, or their lack of prior resources to digitize their collections, for later reuse here, or lack of resources to engage in the entire GA&C process. This hypothesis agrees with our previous results on Russian digital collections (Kizhner et al., 2019) and the findings that few museums across the world pursue open access policies (Fig. 16) (McCarthy and Wallace, 2020). Exclusion bias or difficulty of collecting samples of artworks from places that are difficult to access for a variety of technological, ethical or social reasons (Bode, 2020) can be another reason for bias in cultural data aggregators. Whatever the intentions of Google Cultural Institute\textsuperscript{31} are, and they may, indeed, relate to accessibility and diversity in representing art and culture (Sood, 2016), bias necessarily appears in their corpus in the same way it creeps into linguistic and literary corpora for the reasons of pre-existing social structures, prejudices and canons (Bode, 2020; Caliscan et al., 2017). As a part of a commercial giant, however, GA&C may channel and structure their data to make its algorithms produce more information of similar content (Fylverbom and Murray, 2018; Zuboff, 2019), amplifying existing bias, strengthening prejudices and reinforcing digital cultural colonialism. Unlike a linguistic corpus (Biber, 1993), a scholarly edition or a curated literary system (Bode, 2020), GA&C does not reveal its epistemic choices, limitations and biases.
Fig. 16: Map of the countries with museums that publish their images under open licenses or their analogues, 176 museums from 26 countries, data on museums were extracted from (McCarthy and Wallace, 2020) in February 2019 and adapted as visualization for this paper.

Beyond this discussion of current infrastructure, these biases may also have an effect on future platforms and their users. Existing literature stresses the necessity to have datasets of cultural objects to train convolutional neural networks to get efficient search engines. This means that search engines will be able to answer direct visual queries, determine the subject matter without working with textual metadata and find similar objects in pictures (Di Leonardo et al., 2016; Wevers and Smits, 2019). Such inferences are important as it can contribute to finding a change in patterns across time, genres and geographic regions for studies in art history, historical research and other fields in the humanities (Manovich, 2015) as we move towards thinking of “Collections as Data” (Allen et al., 2019). The pre-existing bias inherited from physical collections of cultural objects that we have identified here may be amplified in visual corpora, resulting in digital cultural colonialism that becomes embedded
into machine-learning protocols and platforms, by the very nature of the data that they are built upon. Our conclusions therefore should not only be taken as a warning for the data structures that we have today, but for the machines that will use them as part of Google’s services, tomorrow.

6. Further work

We hope that our study can motivate practitioners to look at the task of building aggregators in the light of representation and balance. Further work should consider what a representative corpus may be: should we build one on the principles of production (how much art of each type exists in cultural institutions) or on the principles of importance and cultural distinction (or perception)? If the principle of production is important than we should know what types of art, periods and artists dominate the art of the countries an aggregator represents. If this data is not available (as is often the case), data mining can be used to compare national museums and aggregators and how they are clustered. If the principle of importance (or cultural valuing) is relevant than new concerns arise. Further work could look at critical appraisal, consumption rates, and dissemination of artworks with an aim of building a representative corpus as it was previously done for classical music (London, 2013, White and Quinn, 2014) and literature (Bode, 2018). This, in its turn, brings in the question of whether an aggregator should enlarge the canon, and to what extent it should do so. We also hope the paper can generate further work on Google collecting strategies and editorial principles. Future studies will consider how special collections and themes are curated on the GA&C web page, especially its home page. This will generate further studies on how representative these themes are of the aggregator’s content and the content of submitting institutions and submitting countries.

7. Conclusion
Our results clearly demonstrate that GA&C is not a representative and balanced aggregator of digitized cultural content and that it inherits cultural biases of the print era, which results in digital cultural colonialism. This bias is strengthened by the popular character of the aggregator and the dominance of 20th century materials. We find considerable discrepancy in how countries are represented at the level of holding institutions and at the level of foreign objects in other countries’ institutions. We also find the dominance of contemporary paintings and photography in Russian and French collections published by GA&C platform. Our preliminary results based on the analysis of images from Russian and French collections show some emphasis on ‘colonial culture’ phenomenon and the representations of the “exotic other” that can match the expectations of the Western public. While the bias resulting from digitization of historical documents and literary datasets has been discussed in literature (Bode 2018; Zaagsma, 2019; Bode 2020), the magnitude of our results shows the importance of this issue for future digitization projects and aggregators of cultural content. We also find that the difference in the number of objects for different countries and institutions is related to data transparency, open access policies and having previously digitized content to facilitate data transfer. The popular character of the aggregator and the fact that museums sometimes receive recommendations related to the type of content from Google Cultural Institute can be a reason for the dominance of 20th century images.

GA&C is not a scholarly edition and reproduces the bias of the previous era. Data on less visible cultures continue to be distorted and invisible, and cultural layers that are given less prominence are withdrawn. This is in line with recent appeals to decolonize datafication and to correct the injustice of unbalanced data representation (Milan and Treré, 2019). The unbalanced data representation on display here may lead to cultural capital staying in the same ethnic/economic/social groups, which results in decrease in social capital and social mobility, lack of trust and increase in conflicts (Bertrand, 2018). This could be improved through open access policies and the willingness coming from cultural institutions to see their images appearing on external platforms, as well as an attempt by Google to publish its methodologies, selection and epistemic choices plans, and licensing agreements.
(which we have not touched on in this paper beyond the reuse of open licensed content, but questions remain regarding the licensing of content that Google has digitized itself for this project, or the retrospective licensing it may apply for reuse of digitized content (Dewey 2015)). This could be further informed through the extensive discussions in literature in computational literary studies and corpus linguistics on how to address bias related to the lack of representativeness and balance. We could also carry out similar studies on other aggregators, for example, Europeana32, or the commercial Useum33.

It would be naive to think that we can currently avoid exclusion bias, and consequently, digital cultural colonialism, whether at the level of geography, cultural canons or the motives of data appropriation channeled to train algorithms for commercial interests, which results in amplification of prejudices. However, if the equilibrium cannot be found so that cultures are equally and proportionally represented in aggregated digitized content operated by commercial companies and public institutions, we can at least hope to understand ‘the nature of ontological gaps and epistemological biases in its evidence’ (Bode, 2020, p. 2). In this case aggregators can show their limitations within which their validity as corpora can be less questionable (Egbert, 2019). If digitized content analysis produces biased results, is all digitization nothing less than a postcolonial way to exclude cultures from outreach, dissemination and research? Or does this impact come from all parties concerned - social structures, aggregators and memory institutions - and does it have deeper epistemic and social implications? We end with a challenge to Google: what will GA&C do to make its processes for ingesting and showcasing “arts and culture” transparent, and how will it deploy its resources to expand the reach and spread of the digital content featured in its aggregator?

Acknowledgements

The authors are grateful to the DSH anonymous reviewers for their important contributions and comments that significantly improved the paper. We would like to thank Itzhak Benenson and Boris Orekhov for valuable discussions.
Notes


2. https://artsandculture.google.com/?hl=en

3. https://www.google.com/culturalinstitute/about/


5. https://artsandculture.google.com/search?q=kazakhstan

6. https://artsandculture.google.com/category/place

7. The analysis included several stages, with the first stage of looking at the number of countries and collections completed by October 30th, 2018 and the second stage that produced a closer look at French and Russian collections completed by June 15th, 2019.


9. See, for example, https://artsandculture.google.com/partner/mus%C3%A9e-des-beaux-arts-de-lyon


13. https://artsandculture.google.com/entity/russia/m06bnz?categoryId=place, https://artsandculture.google.com/entity/france/m0f8l9e?categoryId=place


15. The National Catalogue of the RF Museum Collections is a comprehensive database of all museums records in the country collected for inventory purposes. By 2026, it is supposed to include 100% of museums objects from all Russian museums reporting to the Ministry of Culture (about 60
million objects). The access to the database is performed through faceted search with two facets, collection types and museums. It can be found online at https://goskatalog.ru/portal/#/. At the time of writing the database includes almost 15 million objects, about 25% of Russian museum collections.

18. https://www.pop.culture.gouv.fr/

19. The list of countries with the number of objects submitted to GA&C is reported in Appendix 1.

20. LIFE magazine was focused on photojournalism, and published weekly from 1936 to 1972. It is a rich source of visual data on American history and social studies (Grady, 2007).

21. As of August 2019, GA&C has new visual collections of the National Aeronautics and Space Administration (NASA) of ca. 160,000 images, which brings North American collections to the top of the list even if we exclude the LIFE Photo Collection.

22. Three exceptions in 2018 were Lithuania, São Tomé and Príncipe, and Equatorial Guinea.

23. https://www.nhm.ac.uk/

27. https://openglam.org

28. A series of ballets staged by Sergei Diaghilev who owned a ballet company and staged ballets in a variety of cities such as Paris and London.

29. These are the museums that are not included in the list of the French Ministry of Culture.

30. La Nouvelle République, a French newspaper, finds some similarity between Porracchia’s drawings and Diaghilev’s Ballets Russes mentioned in Section 3.2 in connection with Russian top creators represented in GA&C (La Nouvelle République, 2018). Cultural transfers showing eastern
influence on the European fashion in the twentieth century are mentioned in a story accompanying Porracchia’s artwork (Imbert, 2019).

31. https://www.google.com/culturalinstitute/about/

32. https://www.europeana.eu/portal/ru

33. https://useum.org
References


Terras, M. 2016. “Crowdsourcing in the Digital Humanities.” In Schreibman S.,


Wevers M., Smits T. (2019). The Visual Digital Turn. Using Neural Networks to Study


Appendix 1

Number of collections, exhibits in collections and number of exhibits in ‘Discover this place’ section as of June 2019 (LIFE Photo Collection was excluded from the analysis).

<table>
<thead>
<tr>
<th>№</th>
<th>Country</th>
<th>Number of Collections</th>
<th>Number of Exhibits (in Collections)</th>
<th>Number of Exhibits (Discover this place)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>USA</td>
<td>413</td>
<td>310 407</td>
<td>1 595 063</td>
</tr>
<tr>
<td>2</td>
<td>Italy</td>
<td>111</td>
<td>98 225</td>
<td>59 755</td>
</tr>
<tr>
<td>3</td>
<td>United Kingdom</td>
<td>108</td>
<td>334 558</td>
<td>79 459</td>
</tr>
<tr>
<td>4</td>
<td>Spain</td>
<td>102</td>
<td>11 985</td>
<td>13 124</td>
</tr>
<tr>
<td>5</td>
<td>Germany</td>
<td>101</td>
<td>18 610</td>
<td>46 111</td>
</tr>
<tr>
<td>6</td>
<td>France</td>
<td>87</td>
<td>16 051</td>
<td>85 390</td>
</tr>
<tr>
<td>7</td>
<td>Japan</td>
<td>81</td>
<td>11 227</td>
<td>66 064</td>
</tr>
<tr>
<td>8</td>
<td>Brazil</td>
<td>59</td>
<td>16 693</td>
<td>11 128</td>
</tr>
<tr>
<td>9</td>
<td>India</td>
<td>55</td>
<td>10 738</td>
<td>30 438</td>
</tr>
<tr>
<td>10</td>
<td>South Korea</td>
<td>54</td>
<td>52 214</td>
<td>38 386</td>
</tr>
<tr>
<td>11</td>
<td>Russia</td>
<td>49</td>
<td>5 520</td>
<td>41 009</td>
</tr>
<tr>
<td>12</td>
<td>Australia</td>
<td>48</td>
<td>13 066</td>
<td>11 438</td>
</tr>
<tr>
<td>13</td>
<td>Mexico</td>
<td>32</td>
<td>4 218</td>
<td>17 312</td>
</tr>
<tr>
<td>14</td>
<td>Canada</td>
<td>30</td>
<td>7 191</td>
<td>17 929</td>
</tr>
<tr>
<td>15</td>
<td>China</td>
<td>29</td>
<td>6 130</td>
<td>35 297</td>
</tr>
<tr>
<td>16</td>
<td>Poland</td>
<td>28</td>
<td>5 431</td>
<td>6 079</td>
</tr>
<tr>
<td>17</td>
<td>Netherlands</td>
<td>27</td>
<td>170 201</td>
<td>14 712</td>
</tr>
<tr>
<td>18</td>
<td>Malta</td>
<td>27</td>
<td>547</td>
<td>454</td>
</tr>
<tr>
<td>19</td>
<td>Austria</td>
<td>18</td>
<td>6 809</td>
<td>8 344</td>
</tr>
<tr>
<td>20</td>
<td>Slovakia</td>
<td>18</td>
<td>1 767</td>
<td>676</td>
</tr>
<tr>
<td>21</td>
<td>Belgium</td>
<td>17</td>
<td>3 534</td>
<td>8 239</td>
</tr>
<tr>
<td>22</td>
<td>Czech Republic</td>
<td>17</td>
<td>2 722</td>
<td>3 478</td>
</tr>
<tr>
<td>23</td>
<td>Portugal</td>
<td>17</td>
<td>2 524</td>
<td>3 048</td>
</tr>
<tr>
<td>24</td>
<td>Hungary</td>
<td>13</td>
<td>1 832</td>
<td>2 873</td>
</tr>
<tr>
<td></td>
<td>Country</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>25</td>
<td>Romania</td>
<td>13</td>
<td>1748</td>
<td>2745</td>
</tr>
<tr>
<td>26</td>
<td>Ireland</td>
<td>12</td>
<td>2454</td>
<td>3624</td>
</tr>
<tr>
<td>27</td>
<td>Republic of South Africa</td>
<td>12</td>
<td>2415</td>
<td>4673</td>
</tr>
<tr>
<td>28</td>
<td>Argentina</td>
<td>12</td>
<td>1910</td>
<td>6808</td>
</tr>
<tr>
<td>29</td>
<td>Denmark</td>
<td>12</td>
<td>1323</td>
<td>3472</td>
</tr>
<tr>
<td>30</td>
<td>Turkey</td>
<td>11</td>
<td>1838</td>
<td>5047</td>
</tr>
<tr>
<td>31</td>
<td>Greece</td>
<td>11</td>
<td>1737</td>
<td>9892</td>
</tr>
<tr>
<td>32</td>
<td>Sweden</td>
<td>11</td>
<td>1191</td>
<td>8220</td>
</tr>
<tr>
<td>33</td>
<td>Switzerland</td>
<td>11</td>
<td>945</td>
<td>10922</td>
</tr>
<tr>
<td>34</td>
<td>Peru</td>
<td>10</td>
<td>1502</td>
<td>3680</td>
</tr>
<tr>
<td>35</td>
<td>Indonesia</td>
<td>9</td>
<td>1844</td>
<td>7939</td>
</tr>
<tr>
<td>36</td>
<td>Pakistan</td>
<td>9</td>
<td>815</td>
<td>4058</td>
</tr>
<tr>
<td>37</td>
<td>Latvia</td>
<td>8</td>
<td>774</td>
<td>316</td>
</tr>
<tr>
<td>38</td>
<td>Philippines</td>
<td>7</td>
<td>1342</td>
<td>16279</td>
</tr>
<tr>
<td>39</td>
<td>Thailand</td>
<td>7</td>
<td>744</td>
<td>9757</td>
</tr>
<tr>
<td>40</td>
<td>Israel</td>
<td>6</td>
<td>1318</td>
<td>8303</td>
</tr>
<tr>
<td>41</td>
<td>Kenya</td>
<td>5</td>
<td>1505</td>
<td>2175</td>
</tr>
<tr>
<td>42</td>
<td>Singapore</td>
<td>5</td>
<td>1012</td>
<td>4413</td>
</tr>
<tr>
<td>43</td>
<td>Colombia</td>
<td>5</td>
<td>798</td>
<td>4493</td>
</tr>
<tr>
<td>44</td>
<td>New Zealand</td>
<td>4</td>
<td>2930</td>
<td>3442</td>
</tr>
<tr>
<td>45</td>
<td>Nigeria</td>
<td>4</td>
<td>1067</td>
<td>3772</td>
</tr>
<tr>
<td>46</td>
<td>Norway</td>
<td>4</td>
<td>418</td>
<td>2257</td>
</tr>
<tr>
<td>47</td>
<td>Azerbaijan</td>
<td>4</td>
<td>333</td>
<td>303</td>
</tr>
<tr>
<td>48</td>
<td>Ukraine</td>
<td>4</td>
<td>283</td>
<td>1340</td>
</tr>
<tr>
<td>49</td>
<td>United Arab Emirates</td>
<td>3</td>
<td>1156</td>
<td>600</td>
</tr>
<tr>
<td>50</td>
<td>Croatia</td>
<td>3</td>
<td>671</td>
<td>1715</td>
</tr>
<tr>
<td>51</td>
<td>Bulgaria</td>
<td>3</td>
<td>542</td>
<td>444</td>
</tr>
<tr>
<td>52</td>
<td>Chile</td>
<td>3</td>
<td>365</td>
<td>4179</td>
</tr>
<tr>
<td>53</td>
<td>Slovenia</td>
<td>3</td>
<td>339</td>
<td>295</td>
</tr>
<tr>
<td>54</td>
<td>Qatar</td>
<td>3</td>
<td>277</td>
<td>26</td>
</tr>
<tr>
<td>55</td>
<td>Finland</td>
<td>3</td>
<td>231</td>
<td>8466</td>
</tr>
<tr>
<td>56</td>
<td>Mongolia</td>
<td>3</td>
<td>217</td>
<td>495</td>
</tr>
<tr>
<td>57</td>
<td>Kuwait</td>
<td>2</td>
<td>447</td>
<td>264</td>
</tr>
<tr>
<td>58</td>
<td>Estonia</td>
<td>2</td>
<td>362</td>
<td>295</td>
</tr>
<tr>
<td>No.</td>
<td>Country</td>
<td>Value</td>
<td>Value</td>
<td>Total</td>
</tr>
<tr>
<td>-----</td>
<td>------------------------------</td>
<td>-------</td>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>59</td>
<td>Serbia</td>
<td>2</td>
<td>128</td>
<td>2 437</td>
</tr>
<tr>
<td>60</td>
<td>Georgia</td>
<td>1</td>
<td>388</td>
<td>375</td>
</tr>
<tr>
<td>61</td>
<td>Ecuador</td>
<td>1</td>
<td>308</td>
<td>717</td>
</tr>
<tr>
<td>62</td>
<td>Jordan</td>
<td>1</td>
<td>141</td>
<td>1 421</td>
</tr>
<tr>
<td>63</td>
<td>Tanzania</td>
<td>1</td>
<td>133</td>
<td>607</td>
</tr>
<tr>
<td>64</td>
<td>Benin</td>
<td>1</td>
<td>130</td>
<td>118</td>
</tr>
<tr>
<td>65</td>
<td>Senegal</td>
<td>1</td>
<td>103</td>
<td>161</td>
</tr>
<tr>
<td>66</td>
<td>Nepal</td>
<td>1</td>
<td>102</td>
<td>1 463</td>
</tr>
<tr>
<td>67</td>
<td>Bosnia and Herzegovina</td>
<td>1</td>
<td>99</td>
<td>179</td>
</tr>
<tr>
<td>68</td>
<td>Armenia</td>
<td>1</td>
<td>94</td>
<td>160</td>
</tr>
<tr>
<td>69</td>
<td>Guatemala</td>
<td>1</td>
<td>93</td>
<td>4 574</td>
</tr>
<tr>
<td>70</td>
<td>Uruguay</td>
<td>1</td>
<td>93</td>
<td>1 902</td>
</tr>
<tr>
<td>71</td>
<td>Egypt</td>
<td>1</td>
<td>64</td>
<td>13 013</td>
</tr>
<tr>
<td>72</td>
<td>Luxembourg</td>
<td>1</td>
<td>54</td>
<td>105</td>
</tr>
<tr>
<td>73</td>
<td>Vietnam</td>
<td>0</td>
<td>0</td>
<td>19 411</td>
</tr>
<tr>
<td>74</td>
<td>Cuba</td>
<td>0</td>
<td>0</td>
<td>10 676</td>
</tr>
<tr>
<td>75</td>
<td>Iran</td>
<td>0</td>
<td>0</td>
<td>8 444</td>
</tr>
<tr>
<td>76</td>
<td>Ethiopia</td>
<td>0</td>
<td>0</td>
<td>5 207</td>
</tr>
<tr>
<td>77</td>
<td>Democratic Republic of the Congo</td>
<td>0</td>
<td>0</td>
<td>5 160</td>
</tr>
<tr>
<td>78</td>
<td>Venezuela</td>
<td>0</td>
<td>0</td>
<td>4 431</td>
</tr>
<tr>
<td>79</td>
<td>Tunisia</td>
<td>0</td>
<td>0</td>
<td>3 930</td>
</tr>
<tr>
<td>80</td>
<td>Jamaica</td>
<td>0</td>
<td>0</td>
<td>3 913</td>
</tr>
<tr>
<td>81</td>
<td>Morocco</td>
<td>0</td>
<td>0</td>
<td>3 896</td>
</tr>
<tr>
<td>82</td>
<td>Lebanon</td>
<td>0</td>
<td>0</td>
<td>3 698</td>
</tr>
<tr>
<td>83</td>
<td>Iraq</td>
<td>0</td>
<td>0</td>
<td>3 202</td>
</tr>
<tr>
<td>84</td>
<td>Laos</td>
<td>0</td>
<td>0</td>
<td>3 148</td>
</tr>
<tr>
<td>85</td>
<td>Myanmar</td>
<td>0</td>
<td>0</td>
<td>2 712</td>
</tr>
<tr>
<td>86</td>
<td>Ghana</td>
<td>0</td>
<td>0</td>
<td>2 505</td>
</tr>
<tr>
<td>87</td>
<td>Haiti</td>
<td>0</td>
<td>0</td>
<td>2 384</td>
</tr>
<tr>
<td>88</td>
<td>Bahamas</td>
<td>0</td>
<td>0</td>
<td>2 346</td>
</tr>
<tr>
<td>89</td>
<td>Panama</td>
<td>0</td>
<td>0</td>
<td>2 329</td>
</tr>
<tr>
<td>90</td>
<td>Dominican Republic</td>
<td>0</td>
<td>0</td>
<td>2 000</td>
</tr>
<tr>
<td>91</td>
<td>Malaysia</td>
<td>0</td>
<td>0</td>
<td>1 955</td>
</tr>
<tr>
<td></td>
<td>Country</td>
<td>0</td>
<td>0</td>
<td>1931</td>
</tr>
<tr>
<td>---</td>
<td>----------------</td>
<td>---</td>
<td>---</td>
<td>------</td>
</tr>
<tr>
<td>92</td>
<td>Syria</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>Cambodia</td>
<td></td>
<td></td>
<td>1 864</td>
</tr>
<tr>
<td>94</td>
<td>Afghanistan</td>
<td></td>
<td></td>
<td>1 838</td>
</tr>
<tr>
<td>95</td>
<td>Bolivia</td>
<td></td>
<td></td>
<td>1 761</td>
</tr>
<tr>
<td>96</td>
<td>North Korea</td>
<td></td>
<td></td>
<td>1 527</td>
</tr>
<tr>
<td>97</td>
<td>Cyprus</td>
<td></td>
<td></td>
<td>1 460</td>
</tr>
<tr>
<td>98</td>
<td>Vatican City</td>
<td></td>
<td></td>
<td>1 442</td>
</tr>
<tr>
<td>99</td>
<td>Sri Lanka</td>
<td></td>
<td></td>
<td>1 383</td>
</tr>
<tr>
<td>100</td>
<td>Fiji</td>
<td></td>
<td></td>
<td>1 357</td>
</tr>
<tr>
<td>101</td>
<td>Saudi Arabia</td>
<td></td>
<td></td>
<td>1 190</td>
</tr>
<tr>
<td>102</td>
<td>Costa Rica</td>
<td></td>
<td></td>
<td>1 026</td>
</tr>
<tr>
<td>103</td>
<td>Papua New Guinea</td>
<td></td>
<td></td>
<td>1 017</td>
</tr>
<tr>
<td>104</td>
<td>Marshall Islands</td>
<td></td>
<td></td>
<td>882</td>
</tr>
<tr>
<td>105</td>
<td>Sudan</td>
<td></td>
<td></td>
<td>857</td>
</tr>
<tr>
<td>106</td>
<td>Monaco</td>
<td></td>
<td></td>
<td>834</td>
</tr>
<tr>
<td>107</td>
<td>Algeria</td>
<td></td>
<td></td>
<td>830</td>
</tr>
<tr>
<td>108</td>
<td>Uzbekistan</td>
<td></td>
<td></td>
<td>753</td>
</tr>
<tr>
<td>109</td>
<td>Uganda</td>
<td></td>
<td></td>
<td>734</td>
</tr>
<tr>
<td>110</td>
<td>Palestine</td>
<td></td>
<td></td>
<td>713</td>
</tr>
<tr>
<td>111</td>
<td>Zambia</td>
<td></td>
<td></td>
<td>704</td>
</tr>
<tr>
<td>112</td>
<td>Madagascar</td>
<td></td>
<td></td>
<td>652</td>
</tr>
<tr>
<td>113</td>
<td>Honduras</td>
<td></td>
<td></td>
<td>644</td>
</tr>
<tr>
<td>114</td>
<td>Paraguay</td>
<td></td>
<td></td>
<td>633</td>
</tr>
<tr>
<td>115</td>
<td>Bahrain</td>
<td></td>
<td></td>
<td>630</td>
</tr>
<tr>
<td>116</td>
<td>Nicaragua</td>
<td></td>
<td></td>
<td>630</td>
</tr>
<tr>
<td>117</td>
<td>Yemen</td>
<td></td>
<td></td>
<td>518</td>
</tr>
<tr>
<td>118</td>
<td>Samoa</td>
<td></td>
<td></td>
<td>501</td>
</tr>
<tr>
<td>119</td>
<td>Tonga</td>
<td></td>
<td></td>
<td>441</td>
</tr>
<tr>
<td>120</td>
<td>Kazakhstan</td>
<td></td>
<td></td>
<td>439</td>
</tr>
<tr>
<td>121</td>
<td>Libya</td>
<td></td>
<td></td>
<td>383</td>
</tr>
<tr>
<td>122</td>
<td>Kiribati</td>
<td></td>
<td></td>
<td>369</td>
</tr>
<tr>
<td>123</td>
<td>Salvador</td>
<td></td>
<td></td>
<td>367</td>
</tr>
<tr>
<td>124</td>
<td>Iceland</td>
<td></td>
<td></td>
<td>351</td>
</tr>
<tr>
<td>125</td>
<td>Suriname</td>
<td></td>
<td></td>
<td>336</td>
</tr>
<tr>
<td></td>
<td>Country</td>
<td>Code1</td>
<td>Code2</td>
<td>Code3</td>
</tr>
<tr>
<td>---</td>
<td>---------------------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>126</td>
<td>Bangladesh</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>127</td>
<td>Zimbabwe</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>128</td>
<td>Micronesia</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>129</td>
<td>Gabon</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>130</td>
<td>Solomon Islands</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>131</td>
<td>Macedonia</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>132</td>
<td>Belarus</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>133</td>
<td>Barbados</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>134</td>
<td>Trinidad and Tobago</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>135</td>
<td>Albania</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>136</td>
<td>Dominica</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>137</td>
<td>Tuvalu</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>138</td>
<td>Cameroon</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>139</td>
<td>Sierra Leone</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>140</td>
<td>Rwanda</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>141</td>
<td>Botswana</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>142</td>
<td>Guyana</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>143</td>
<td>Angola</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>144</td>
<td>Mozambique</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>145</td>
<td>Mali</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>146</td>
<td>Palau</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>147</td>
<td>South Sudan</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>148</td>
<td>Somalia</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>149</td>
<td>Andorra</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>150</td>
<td>Moldova</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>151</td>
<td>Namibia</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>152</td>
<td>Saint Kitts and Nevis</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>153</td>
<td>Liberia</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>154</td>
<td>Cote d'Ivoire</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>155</td>
<td>Eritrea</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>156</td>
<td>Antigua and Barbuda</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>157</td>
<td>Burkina Faso</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>158</td>
<td>Eswatini</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>159</td>
<td>Malawi</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Country</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>160</td>
<td>Central African Republic</td>
<td>0</td>
<td>0</td>
<td>58</td>
</tr>
<tr>
<td>161</td>
<td>Montenegro</td>
<td>0</td>
<td>0</td>
<td>57</td>
</tr>
<tr>
<td>162</td>
<td>Belize</td>
<td>0</td>
<td>0</td>
<td>54</td>
</tr>
<tr>
<td>163</td>
<td>Mauritania</td>
<td>0</td>
<td>0</td>
<td>53</td>
</tr>
<tr>
<td>164</td>
<td>Mauritius</td>
<td>0</td>
<td>0</td>
<td>52</td>
</tr>
<tr>
<td>165</td>
<td>Grenada</td>
<td>0</td>
<td>0</td>
<td>49</td>
</tr>
<tr>
<td>166</td>
<td>Niger</td>
<td>0</td>
<td>0</td>
<td>48</td>
</tr>
<tr>
<td>167</td>
<td>Lesotho</td>
<td>0</td>
<td>0</td>
<td>46</td>
</tr>
<tr>
<td>168</td>
<td>Guinea</td>
<td>0</td>
<td>0</td>
<td>45</td>
</tr>
<tr>
<td>169</td>
<td>Seychelles</td>
<td>0</td>
<td>0</td>
<td>42</td>
</tr>
<tr>
<td>170</td>
<td>Republic of the Congo</td>
<td>0</td>
<td>0</td>
<td>36</td>
</tr>
<tr>
<td>171</td>
<td>Turkmenistan</td>
<td>0</td>
<td>0</td>
<td>34</td>
</tr>
<tr>
<td>172</td>
<td>Oman</td>
<td>0</td>
<td>0</td>
<td>32</td>
</tr>
<tr>
<td>173</td>
<td>Brunei Darussalam</td>
<td>0</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>174</td>
<td>Saint Lucia</td>
<td>0</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td>175</td>
<td>Kyrgyzstan</td>
<td>0</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>176</td>
<td>Saint Vincent and the Grenadines</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>177</td>
<td>San Marino</td>
<td>0</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>178</td>
<td>Comoro Islands</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>179</td>
<td>Cabo Verde</td>
<td>0</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>180</td>
<td>Liechtenstein</td>
<td>0</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>181</td>
<td>Tajikistan</td>
<td>0</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>182</td>
<td>Bhutan</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>183</td>
<td>Gambia</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>184</td>
<td>Maldives</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>185</td>
<td>Vanuatu</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>186</td>
<td>Djibouti</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>187</td>
<td>Togo</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>188</td>
<td>Nauru</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>189</td>
<td>Burundi</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>190</td>
<td>Guinea-Bissau</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>191</td>
<td>Chad</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>192</td>
<td>East Timor</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Lithuania</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>----------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>193</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>194</td>
<td>São Tomé and Príncipe</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>195</td>
<td>Equatorial Guinea</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>