

Распределение визуального контента в цифровых библиотеках (на примере цифровой библиотеки «Европеана» и «Мировой цифровой библиотеки»)

Кижнер И.А., Рудов, И.Н. Румянцев М.В., Хохлова В.Н., Афанасьева Ю. Ю.
Сибирский федеральный университет, Научная библиотека,
Красноярск, Россия

Мотивация

- Представление о том, что цифровые библиотеки организуют оцифрованный контент визуального содержания таким образом, чтобы исследователи, которые работают с большими данными получили объективную картину.
- Представление о том, что поскольку мы имеем дело с большим количеством данных, работа по распространению знаний исторического, социального и культурологического характера становится значительно проще и дает дополнительные аналитические возможности определения тематики изображений для дальнейших сравнительных исследований.

Вопросы исследования

- Насколько репрезентативен и сбалансирован визуальный контент цифровых библиотек?
- Получат ли пользователи равное представление о разных культурах?
- Могут ли будущие исследования, основанные на больших данных, продемонстрировать безошибочные и беспристрастные результаты при оценке содержания контента регионов и при сравнении разных культур?

Цель

Оценить репрезентативность и сбалансированность визуального контента, представленного «Европеаной» и «Мировой цифровой библиотекой» и сравнить полученные данные с результатами предыдущего исследования, проведенного на примере цифровой издательской платформы Google Arts and Culture

Европеана



Коллекции ▾ Просмотр ▾ Выставки ▾ Блог ▾

НАШИ САЙТЫ

ЯЗЫК ▾

Добавить поисковый термин



Просмотр

Исследуйте 57,583,288 произведений искусства, предметов материальной культуры, книг, видео и звукозаписей со всей Европы

Ehtoja luetaan, Carl Larsson
Finnish National Gallery

© CC0

СВЯЗЬ

Мировая цифровая библиотека (МЦБ)



МИРОВАЯ
ЦИФРОВАЯ БИБЛИОТЕКА

Просмотр ▾

Временные шкалы

Интерактивные карты

Темы ▾

LIBRARY OF
CONGRESS



При поддержке Организации
Объединенных Наций по
вопросам образования, науки
и культуры

Поиск 19 147 объектов о 193 странах между 8000 до н.э. и 2000:

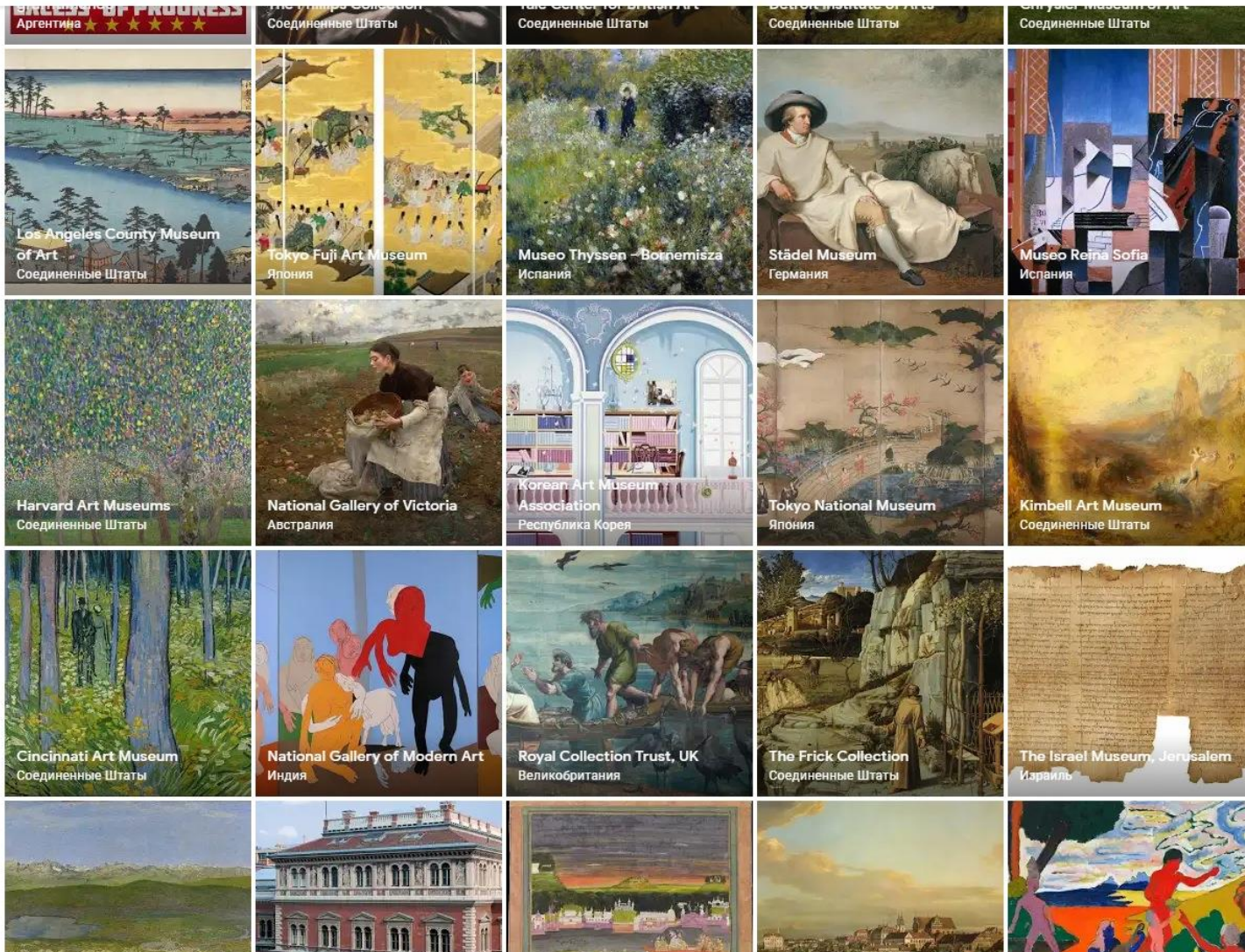
например лошади, исламская архитектура, железные дороги

Поиск

Представленные объекты >



Google Arts and Culture (GA&C)



Методы

«Европеана» и МЦБ предоставляют данные относительно количества учетных записей по странам и типам визуального контента.

ИНСТРУМЕНТЫ ПОИСКА

КОЛЛЕКЦИИ ▲

- All Items
- 1914-1918
- Archaeology
- Art
- Fashion

Больше ▾

НОСИТЕЛЬ ▼

- Только со ссылками на носители

МОГУ ЛИ Я ЭТО ИСПОЛЬЗОВАТЬ? [?](#)

СТРАНА ▼

Place

- Europe 7,276
- Middle East and North Africa 3,629

Time Period

- 1900 - 1949 7,489
- 1850 - 1899 6,464

Topic

- History & geography 16,135
- Computer science, informatio 3,560

Additional Subjects

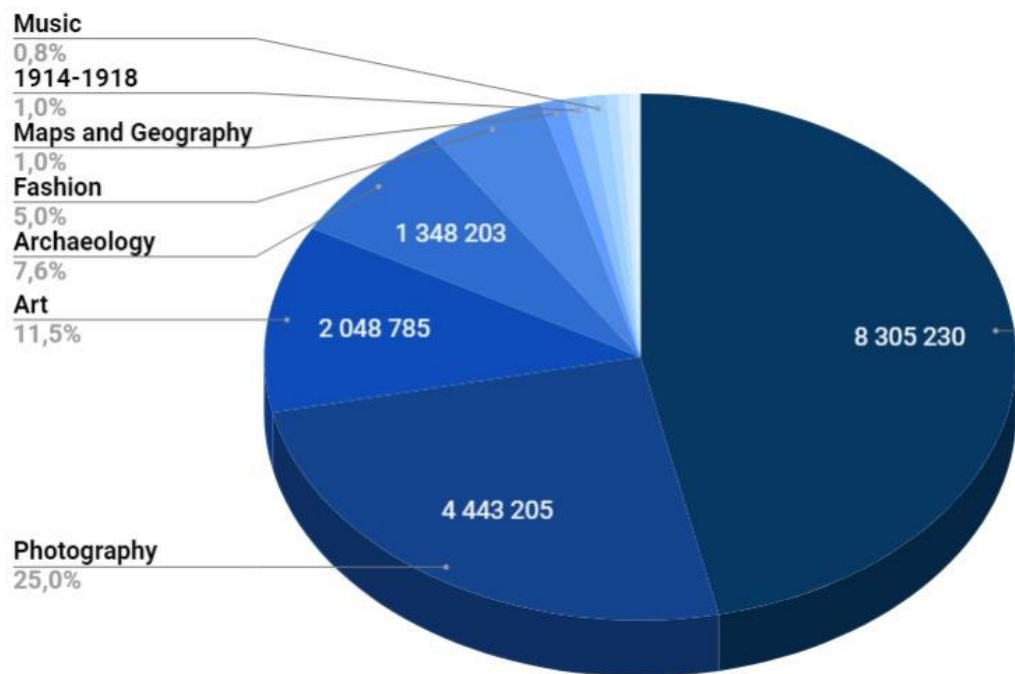
- Meeting of Frontiers 5,000

Методы

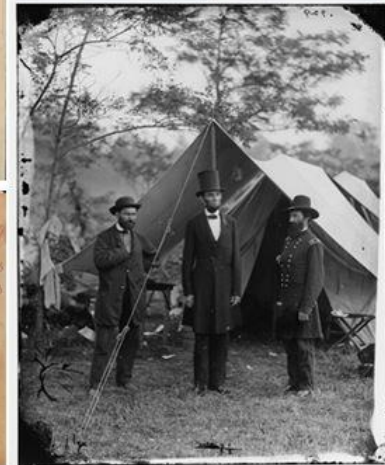
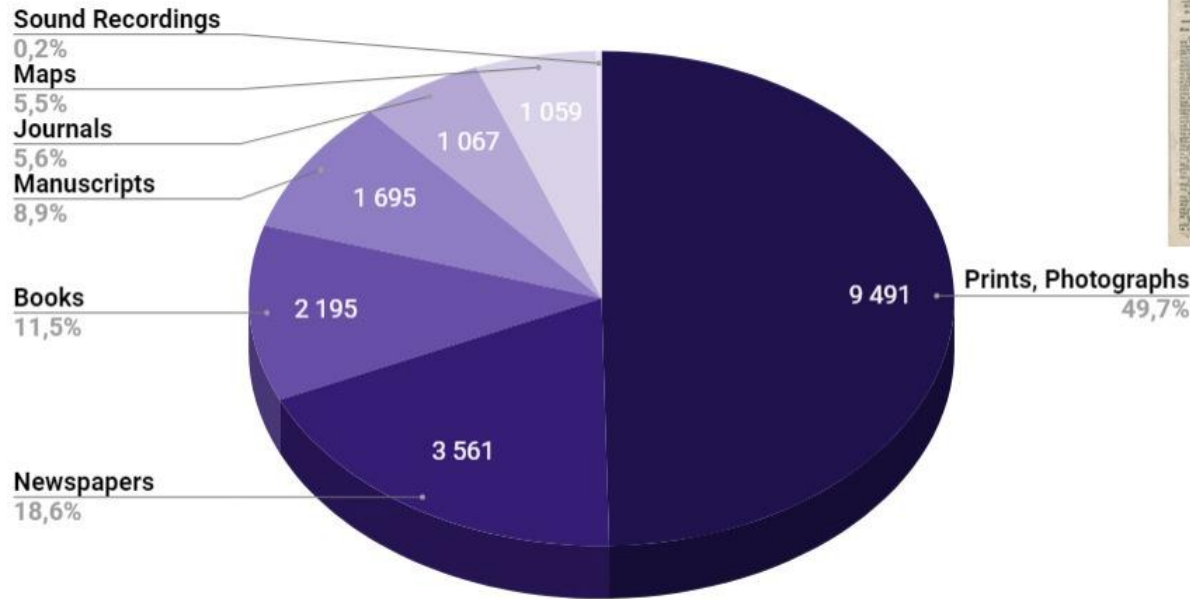
На главной странице сайта GA&C, через опцию «Места» получен доступ к цифровым коллекциям стран. Каждая из стран представлена в проекте GA&C своей отдельной страницей. На странице страны опубликованы цифровые коллекции и материалы по этому месту.

Мы внесли в таблицу эти данные и упорядочили их для того, чтобы выявить распределение контента по странам, периодам и типам изображений.

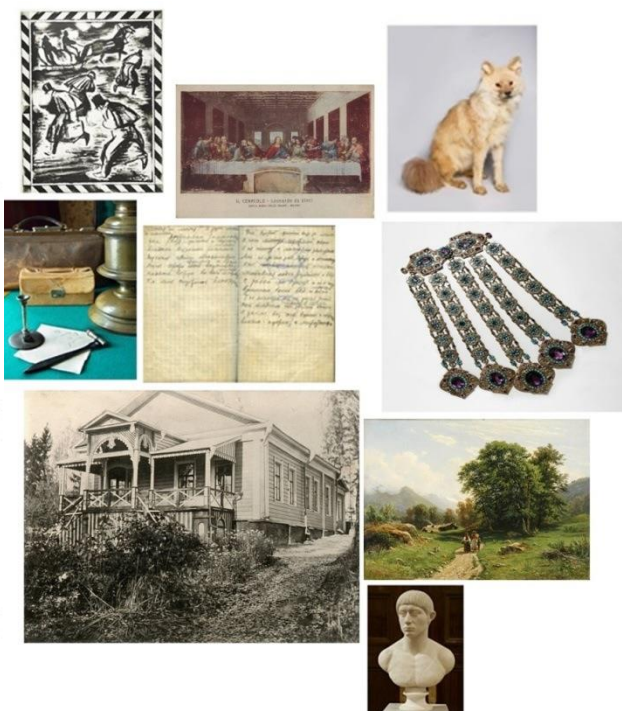
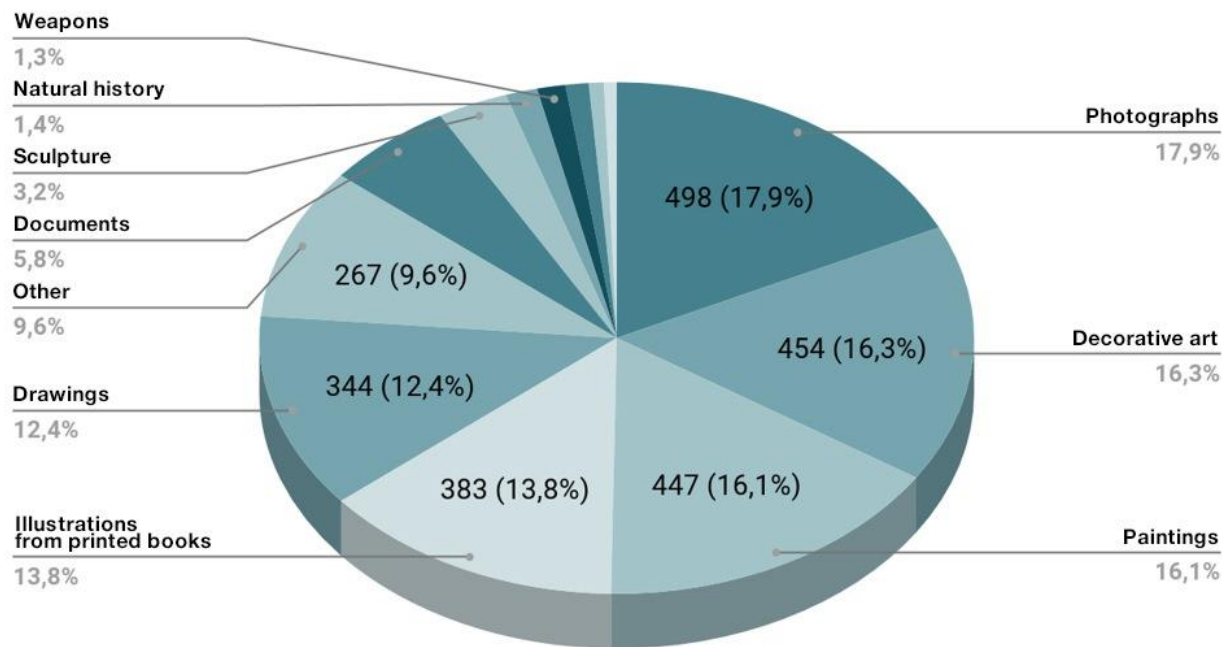
Распределение визуального контента в «Европеане» по типу изображений



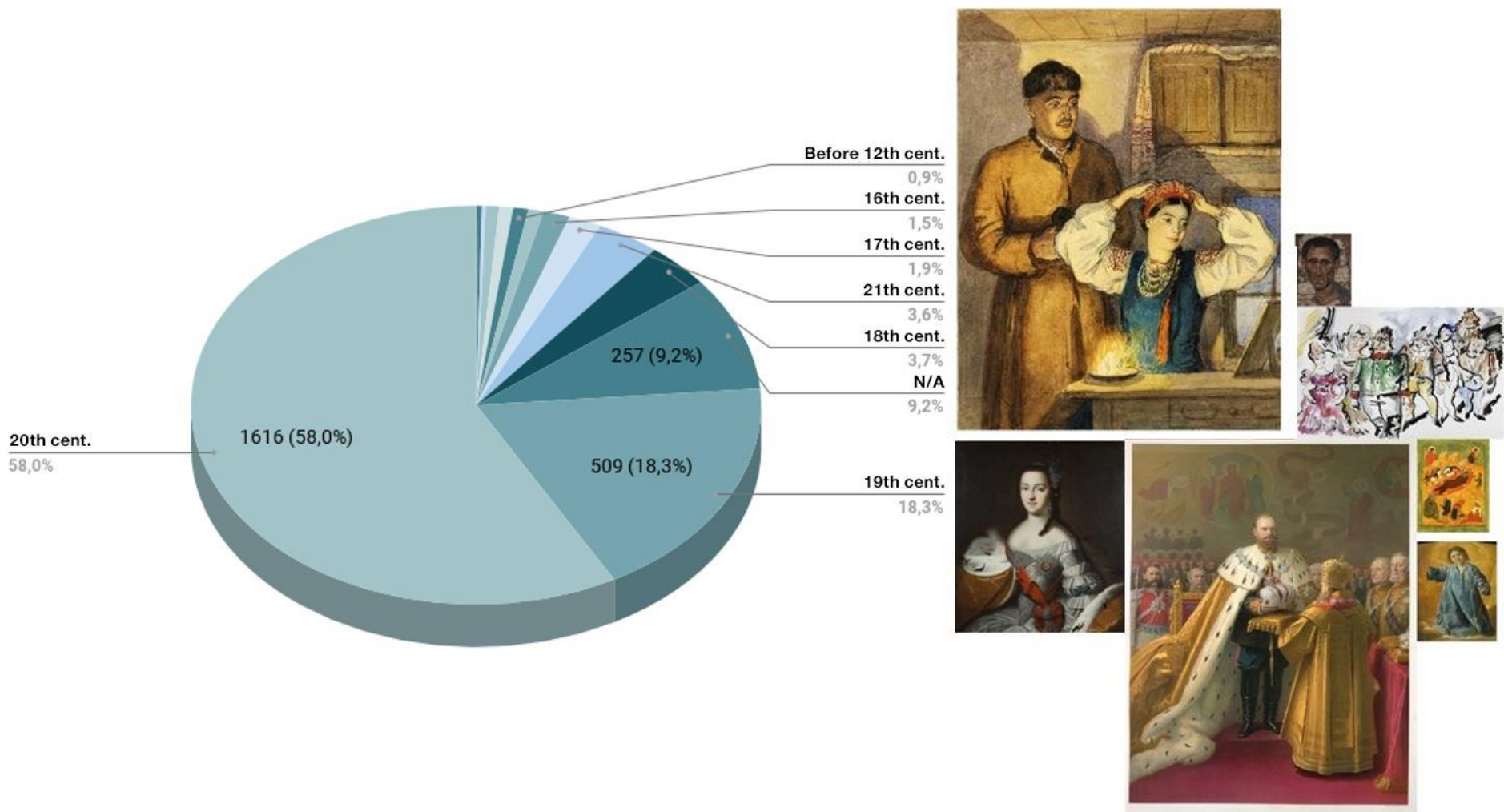
Распределение визуального контента в «Мировой цифровой библиотеке» по типу изображений



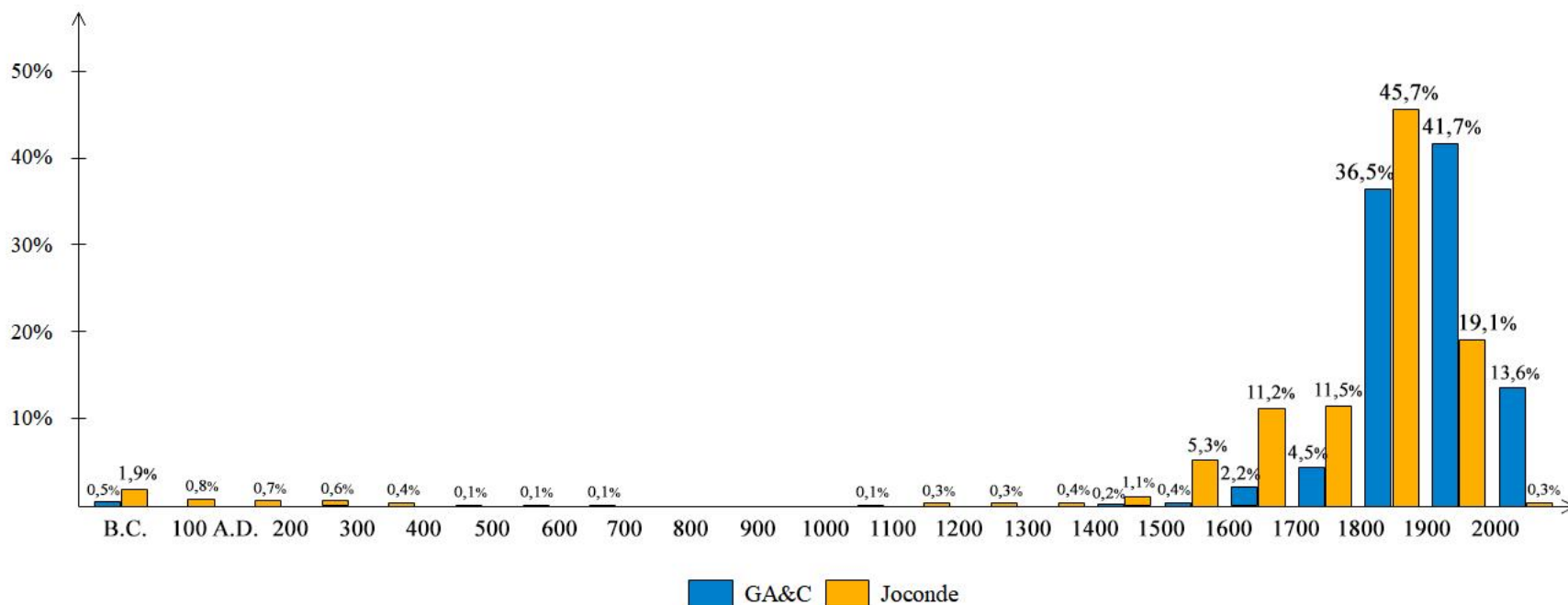
Распределение визуального контента в Google Arts & Culture по типу изображений (на примере коллекций России)



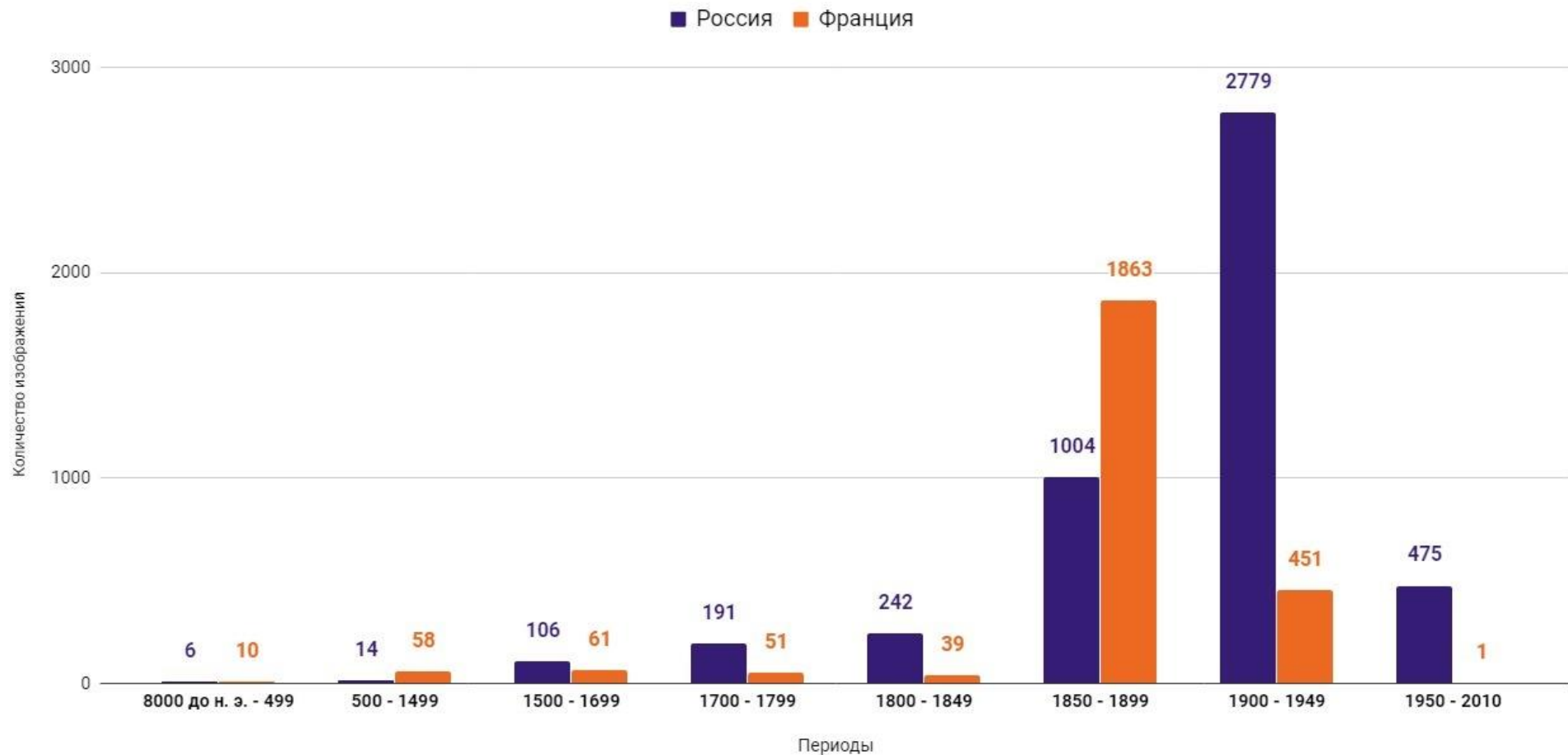
Распределение визуального контента в Google Arts & Culture по периодам (на примере коллекций России)



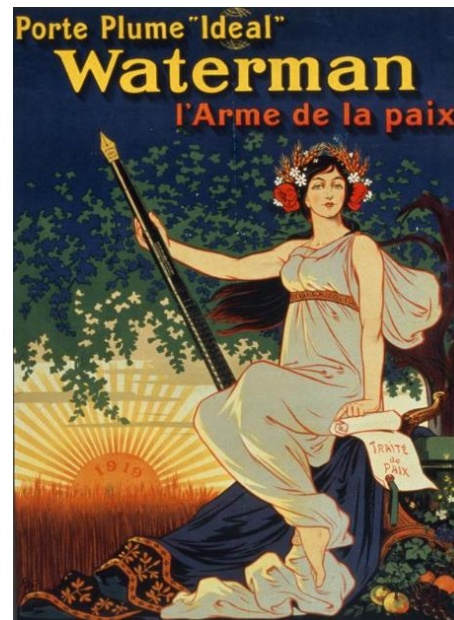
Распределение визуального контента в Google Arts & Culture по периодам (на примере коллекций Франции)



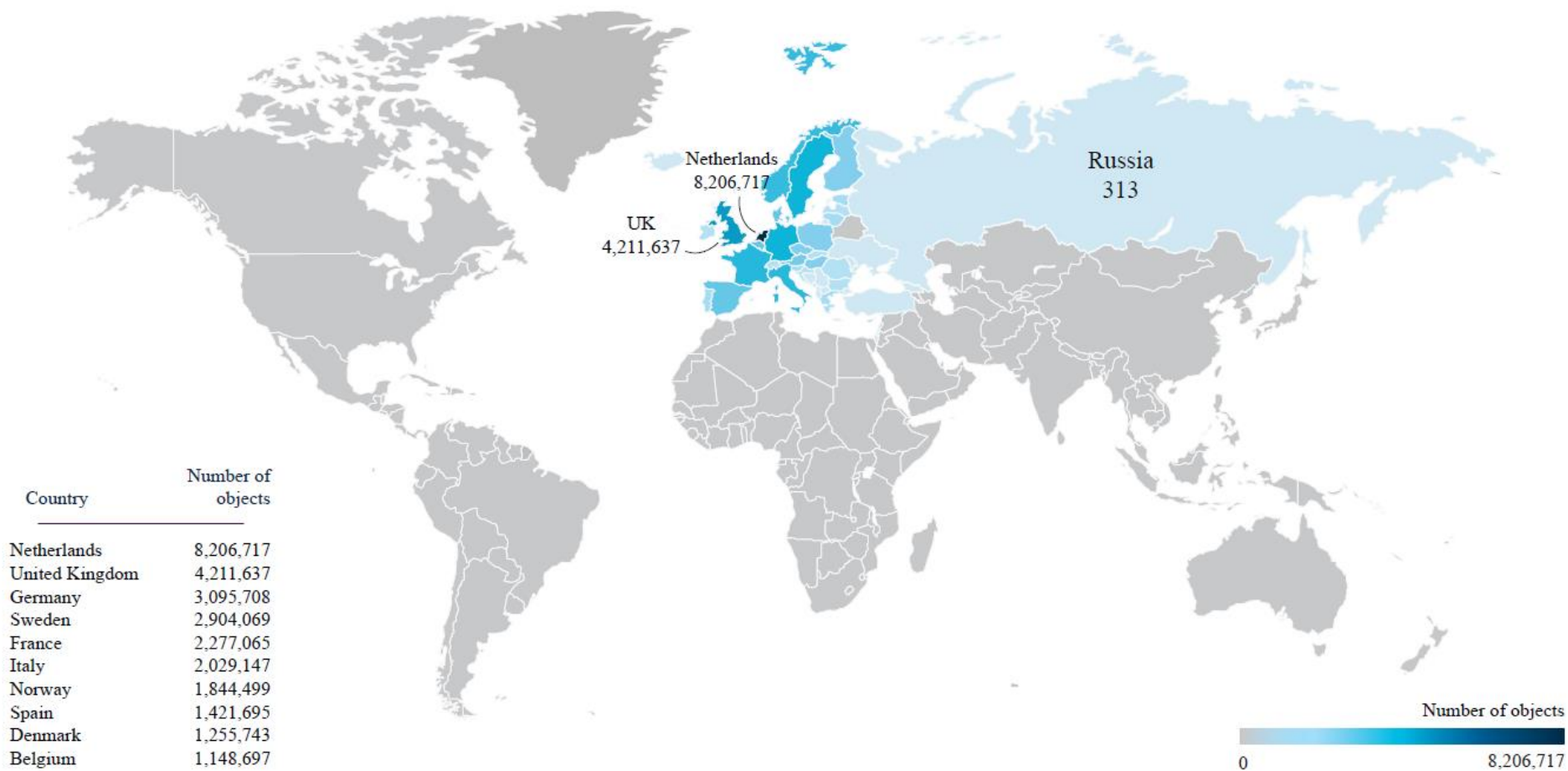
Распределение визуального контента в «Мировой цифровой библиотеке» по периодам (на примере коллекций России и Франции)



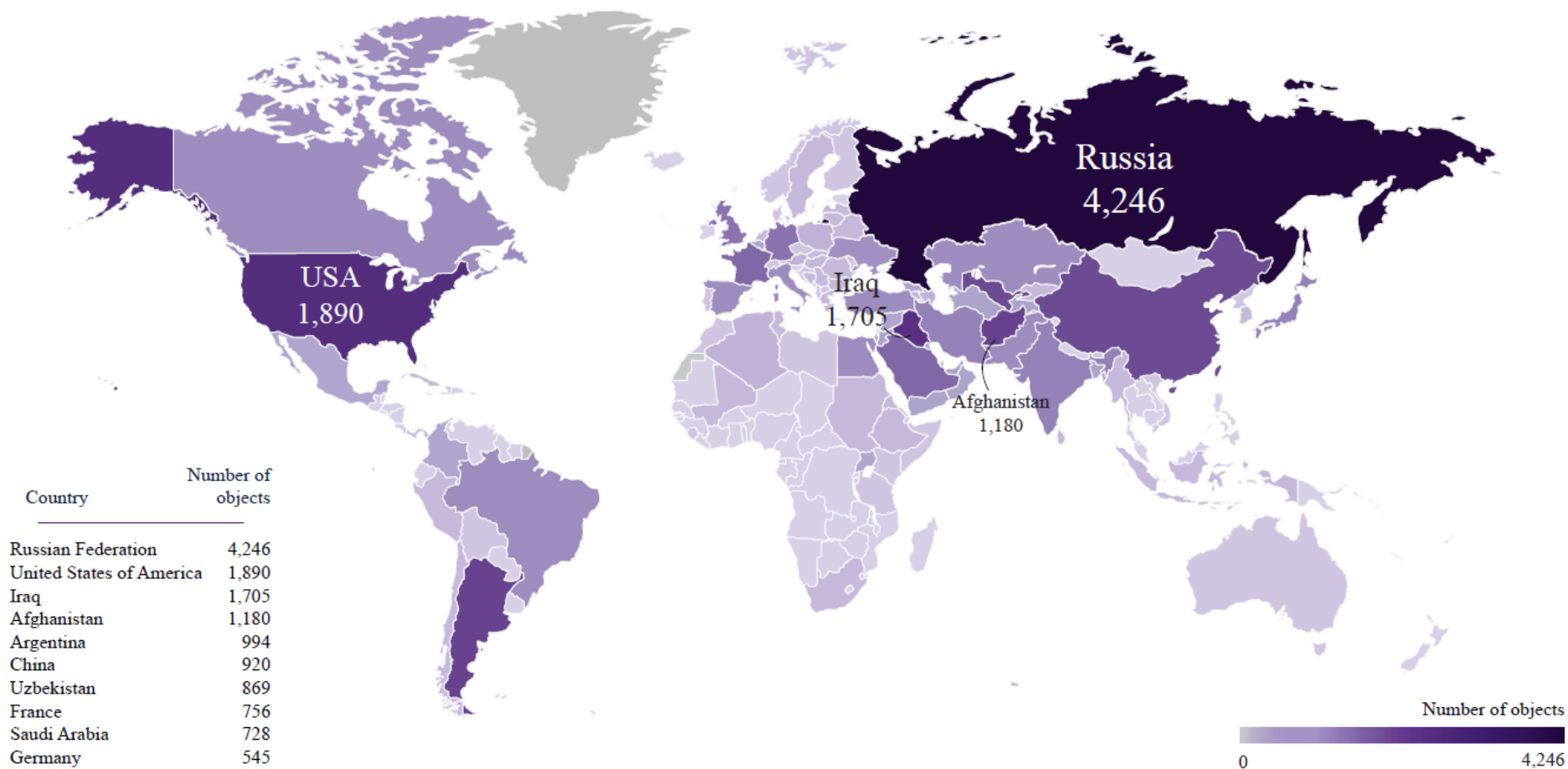
Распределение визуального контента в «Мировой цифровой библиотеке» по периодам (на примере коллекций России и Франции)



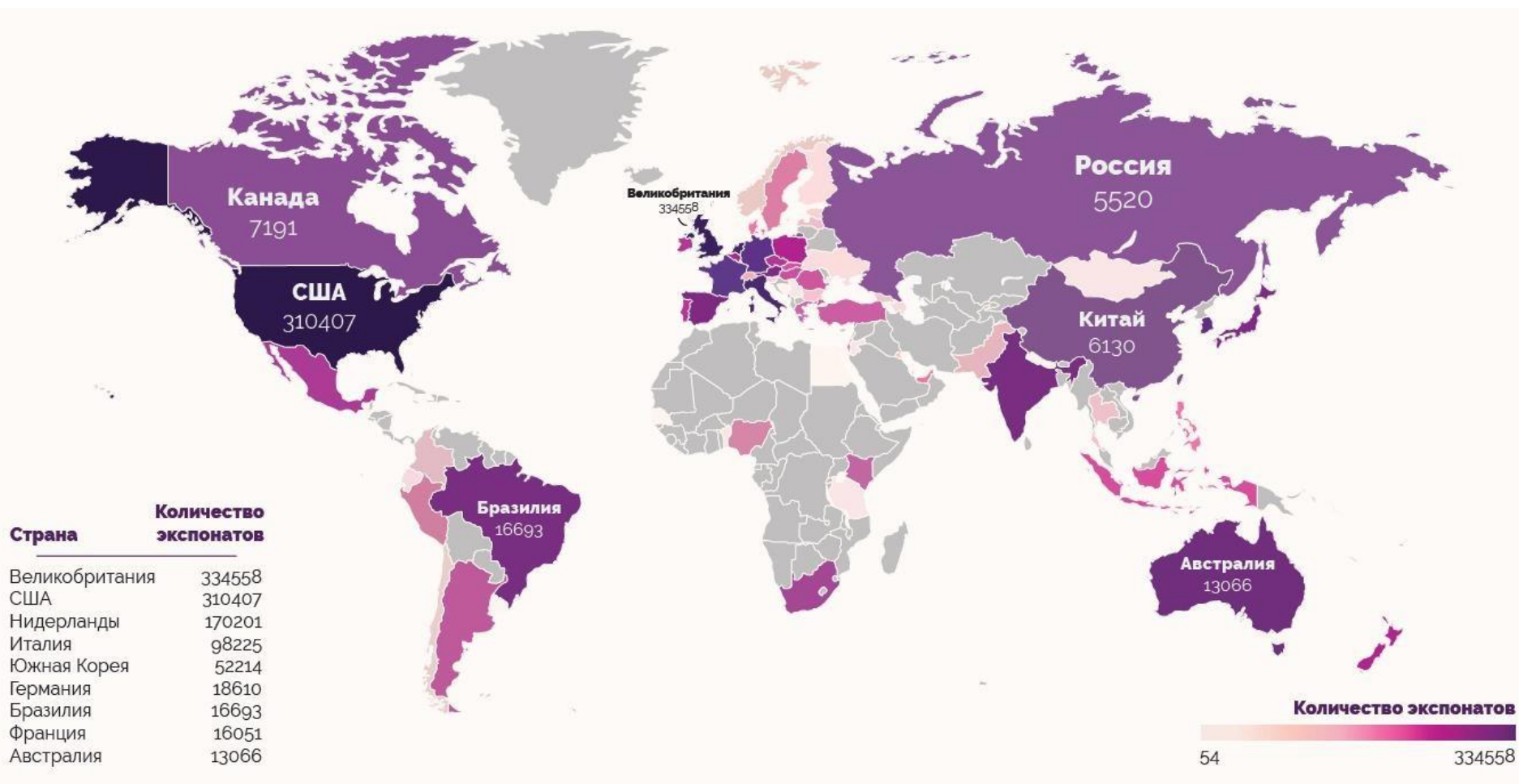
Географическое распределение визуального контента в «Европеане»



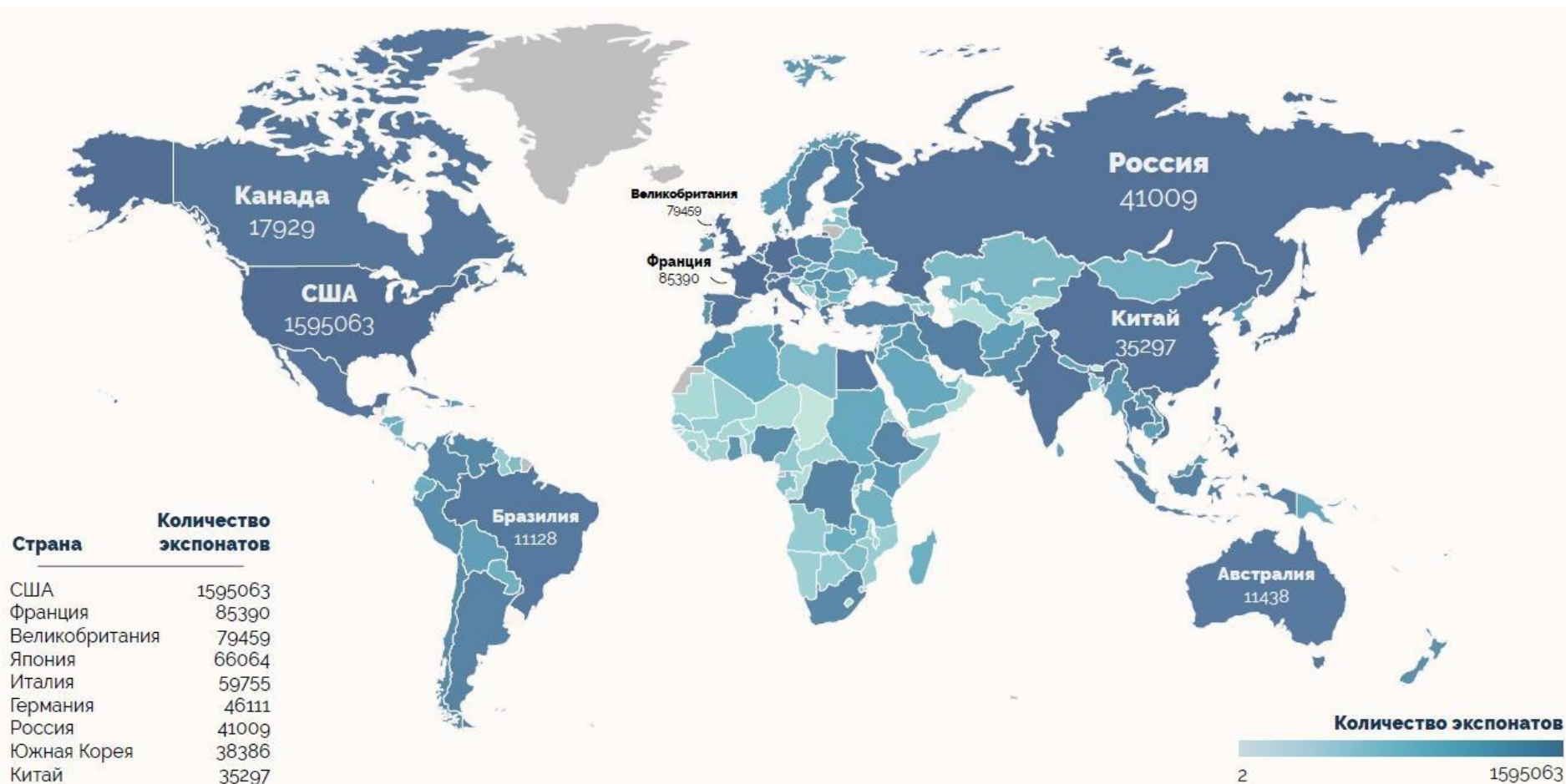
Географическое распределение визуального контента в «Мировой цифровой библиотеке»



Географическое распределение визуального контента в Google Arts & Culture (страны с наибольшим числом экспонатов в коллекциях, не включая крупнейшую коллекцию журнала LIFE «LIFE Photo Collection»)



Географическое распределение визуального контента в Google Arts & Culture (страны с наибольшим числом экспонатов из «материалов по этому месту»)



Результаты

При сравнении результатов полученных для «Европейцы» и МЦБ, мы можем наблюдать близкие тенденции:

- доминирование материалов, которые относятся к XIX-XX векам;
- неравномерность географического распределения материалов.

Географическое распределение материалов для МЦБ демонстрирует большую равномерность и сбалансированность с точки зрения представленности частей света среди стран, опубликовавших наибольшее число материалов в библиотеке, по сравнению с проектом Google Arts and Culture

Результаты

Мы можем констатировать неравномерность и не сбалансированность содержания цифровых библиотек в исследовании.

- анализ больших данных с использованием материалов цифровых библиотек может привести к ошибочным результатам;
- пользователи увидят большее количество материалов, представляющих доминирующие страны и периоды.

Заключение

Цифровой канон будет расширен за счет ряда культур и периодов. Остальные периоды и культуры будут исключены из анализа.

Несмотря на то, что мы видим большую сбалансированность в представлении культур и отдельных регионов в МЦБ по сравнению с платформой Google Arts and Culture, основная часть материалов относится к колониальному и постколониальному периодам и не дает представления о периодах расцвета и независимого развития этих регионов.

Заключение

Было бы наивно предположить, что тенденциозность, связанная с недостаточной представленностью периодов и культур в цифровых библиотеках, агрегирующих визуальный контент, может быть устранена в ближайшие годы. Однако, мы можем надеяться определить, как возникли существующие пробелы и в чем заключаются их социальные, психологические и организационные причины (Bode, 2020).

Заключение

Если анализ оцифрованных данных приведет к появлению тенденциозных результатов, то не является ли оцифровка ничем иным, как постколониальным исключением периодов и культур из распространения и научных исследований? Для объяснения причин этого явления, вероятно, нужно обратиться к тому, как устроено общество, агрегаторы цифрового контента, библиотеки, музеи и архивы. Вероятно, у этого явления есть более глубокие социальные и эпистемические причины.