

Федеральное государственное автономное
образовательное учреждение высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Гуманитарный институт
Кафедра информационных технологий в креативных и культурных
индустриях

УТВЕРЖДАЮ
Заведующий кафедрой
_____ А. В. Усачёв
подпись
« _____ » _____ 2022 г.

БАКАЛАВРСКАЯ РАБОТА

09.03.03.14 Прикладная информатика в области искусств и гуманитарных
наук
Оценка репрезентативности тематических коллекций цифровой библиотеки
Европеана

Руководитель _____ проф., д-р филос. наук Р.П. Мусат
подпись, дата

Выпускник _____ А.П. Орлова
подпись, дата

Красноярск 2022

СОДЕРЖАНИЕ

Введение.....	3
1 Цифровые библиотеки и коллекции.....	7
1.1 Культурная аналитика и цифровые библиотечные системы.....	7
1.2 Современные электронные библиотеки	11
1.3 Цифровая библиотека Европеана	13
2 Материалы и методы исследования.....	16
2.1 Методы исследования.....	16
2.2 Сбор и обработка данных.....	18
2.2.1 Разнообразие контента	21
2.2.2 Сбалансированность данных	27
2.3 Визуализация данных исследования.....	31
3 Результаты исследования.....	33
3.1 Анализ данных по разнообразию материалов коллекций Европеаны ...	33
3.2 Анализ данных по сбалансированности материалов коллекций Еuropeаны	40
3.3 Обсуждение результатов исследования.....	44
Заключение	46
Список использованных источников	48

ВВЕДЕНИЕ

Цифровые коллекции культурного наследия и их агрегаторы предоставляют большое количество инструментов и материалов для проведения исследований в различных направлениях цифровых гуманитарных наук. С целью проведения классификации собранных материалов агрегаторы предоставляют возможность фильтрации содержимого коллекций для проведения более точного поиска и перемещению по ним. Однако разнообразие категорий фильтрации часто является следствием кураторского отбора, а не подбором репрезентативной выборки данных. Недостаточная представленность или полное отсутствие представлений о культурах и культурных сообществах в агрегаторах данных напрямую влияет на формирование культурно-исторического капитала или степени информированности о иных культурах и представлений о культурной идентичности других сообществ.

Смещение точки фокуса с одних культур на другие может проявляться в типе контента, также объекты коллекций могут быть смещены в стороны одного учреждения культуры или нескольких крупных стран. Такой скос в данных может быть скрыт за интерфейсом, который создает впечатление разнородности тематик, отображает межкультурные отношения, создает образ мультикультурного пространства.

Выпускная квалификационная работа посвящена количественному анализу репрезентативности, сбалансированности и разнообразия тематических коллекций европейской цифровой библиотеки Европеана.

Достоверность результатов определяется выборкой данных, включающей 260 записей о культурных объектах из тематических коллекций представленных в главном интерфейсе библиотеки Европеаны. Эти записи составляют 23,5 % от общего количества 1107 записей по 4 разделам, которые расположены на главной странице интерфейса. Также были собраны

данные о количестве учетных записей по 197 странам, общий объем по которым составил 39 094 452 учетные записи на начало декабря 2021 года.

Цифровая библиотека историко-культурного наследия Европеана может выступать в качестве первичного источника данных для культурологических и гуманитарных исследований при условии репрезентативности, сбалансированности и разнообразия размещаемого в тематических коллекциях контента.

Проблема исследовательской работы заключалась в том, что библиотека Европеана не может выступать первичным источником в культурологических и гуманитарных исследованиях. Связано это с тем, что не было проведено исчерпывающее исследование для установления уровня репрезентативности, сбалансированности и разнообразия контента тематических коллекций.

Актуальность настоящей работы заключается в том, что агрегаторы цифровых коллекций в области культурно-исторического наследия всё чаще становятся способами познания мира, а также первостепенными источниками для большого количества гуманитарных исследований, вследствие чего образуется необходимость в создании ресурсов наполненных разнообразным, репрезентативным и сбалансированным контентом, который создаст целостное мультикультурное пространство для исследователей и рядовых пользователей.

Объект исследования - тематические коллекции цифровой библиотеки Европеана.

Предметом исследования представляется сбалансированность и разнообразие как часть понятия репрезентативности тематических коллекций цифровой библиотеки Европеана.

Цель настоящей работы - определить, является ли цифровая библиотека Европеана сбалансированным и разнообразным источником данных для гуманитарных исследований, используя количественный метод анализа.

Задачи исследования:

- обозначить понятие репрезентативности и сбалансированности, изучить особенности цифровых библиотечных систем и репрезентативных выборок для музейных коллекций;
- собрать данные о цифровых коллекциях и интерфейсе Europeana;
- выявить и проанализировать количественные особенности полученных данных;
- визуализировать полученные данные в виде диаграмм;
- провести анализ, определить разнообразие коллекций и снос данных в географических метаданных Europeana и объяснить полученные результаты.

Научная новизна настоящего исследования обосновывается малочисленными исследованиями агрегаторов и цифровых библиотек с точки зрения определения их места в развитии цифровой культуры и применением нового подхода в установлении репрезентативности цифровых коллекций для культурологических и гуманитарных исследований. Новый подход в изучении больших тематических подборок данных основывается на методах заимствованных из корпусной лингвистики и предполагает количественный анализ данных.

Выпускная квалификационная работа состоит из введения, трёх глав, заключения, списка использованной литературы и приложения.

Первая глава «Цифровые библиотеки и коллекции» описывает и раскрывает понятие репрезентативности и сбалансированности, содержит информацию об особенностях проекта Europeana.

Вторая глава «Материалы и методы исследования» описывает методы, применяемые в процессе текущей исследовательской работы.

Третья глава «Результаты исследования» включает описание и обсуждение полученных в ходе исследования результатов анализа тематических цифровых коллекций Europeana.

В заключении подводятся итоги проведенного исследования и его результатов. Список литературы представляет библиографическое описание цитируемых в работе источников. В приложении представлен список государств-членов ООН, который использовался для оценки географического распределения оцифрованного контента.

1 Цифровые библиотеки и коллекции

1.1 Культурная аналитика и цифровые библиотечные системы

За последние 30 лет оцифровка культурного наследия открыла много новых возможностей в изучении культурного прошлого с использованием методов компьютерного анализа больших данных. Сегодня вычислительные методы применяются во многих областях науки, например, в информатике, вычислительной социологии, истории цифрового искусства, цифровых гуманитарных науках и культурной аналитике. Еще в 2007 году Лев Манович ввел термин “культурная аналитика” для обозначения исследований, связанных с визуальными цифровыми коллекциями, независимо от того какие аналитические методы будут применены. Цель этого подхода в том, чтобы дать возможность пользователям осуществлять “глубокий культурный поиск” с помощью создания инструментов с открытым исходным кодом. Это позволит проводить обработку и сортировку данных нестандартными способами [1; 2].

Формирование цифровых тематических коллекций культурно-исторического наследия народов мира уже стало новой вехой профессионального взаимодействия международного библиотечного сообщества [3]. Экспансия цифровых технологий ведет к крайне быстрому рождению новых информационных потоков, которые становятся объектами цифровых гуманитарных наук. Вдобавок, исследование новых источников информации сопровождается увеличению их числа. К ним относятся как первичные источники, формируемые эпохой и людьми, так и вторичные, которые создаются в процессе применения различных методов исследования [4].

Для того чтобы продемонстрировать ограничения электронных массивов агрегированного визуального контента, применяется понятие

репрезентативных и сбалансированных цифровых коллекций. Количественный метод анализа состава коллекции для оценки репрезентативности и сбалансированности цифровых коллекций заимствован из первых исследований цифровых коллекций лингвистических данных. Разработанные в корпусной лингвистике принципы повлияли на формирование и методики изучения больших массивов информации, особенно характерных для гуманитарных областей наук. Первые работы по индексации и документированию объектов культурного наследия для машинной обработки затрагивали только каталогизацию коллекций и увеличения технических возможностей цифровых библиотек культурного наследия, как следствие это привело к увеличению их объемов и масштабов взаимодействия с пользователями [5].

Не существует единого общепринятого определения термина цифровой библиотеки, но все они во многом схожи. В англоговорящих странах, откуда и появился этот термин, принято говорить «цифровая библиотека¹», однако в русскоязычных информационных источниках принято употреблять словосочетание «электронная библиотека», но лексическое значение данного словосочетания при этом не изменяется [6].

В контексте данного исследования под цифровой библиотекой будет пониматься библиотека, имеющая оцифрованные или электронные источники информации, являющаяся развивающейся корпоративной экосистемой со сложной разметкой данных, доступ к которой обеспечивается посредством электронных устройств [7; 8]. Марчук А. Г. и Марчук П.А. утверждают, что чаще всего цифровая библиотека представляет собой множество метаинформационных записей об объектах, находящихся в фонде библиотеки. Сами объекты могут существовать в виде электронного образа² и являться важной частью библиотечной системы [9].

¹ Digital Library

² оцифрованные фотоснимки, видеозаписи, картины, сканы документов, 3D модели археологических находок и др.

В текущем исследовании будут рассмотрены некоторые термины, понятия и их свойства, заимствованные цифровыми гуманитарными науками из компьютерной лингвистики. К таким перекочевавшим терминам относятся лингвистические текстовые корпуса, изучением которых занимается корпусная лингвистика.

Термин «корпус текстов» в лингвистике обладает выраженной полисемией. Но в корпусной лингвистике определение этого понятия принято основывать на четырех основных признаках, а именно:

1. расположение корпуса на машинном носителе;
2. стандартизированность представления данных, что позволяет применять стандартные программы для их обработки;
3. конечный размер;
4. репрезентативность как результат особой процедуры отбора данных.

Первые три признака определяют некий набор текстовых данных, размещаемых на цифровом носителе и представленных в удобном для программной обработки виде. Последний признак репрезентативности подразумевает, что корпус должен максимально объективно представлять разнообразие изучаемого явления, и в тоже время давать объективную картину бытования этого явления в практике [10; 11].

Другие исследователи утверждают, что главной отличительной чертой корпуса является обработка текста специально разработанной машинной разметкой. Тогда собрание текстов в электронной форме не обладающих разметкой на специально созданном техническом языке правильно следует называть электронной или цифровой библиотекой. Электронные библиотеки призваны предоставить доступ пользователям для прочтения текстов. Хотя корпус представляет собой крупное собрание текстов, используется он не для чтения [12; 13].

В свою очередь, еще одним важным свойством корпуса является его сбалансированность, которая позволяет определить равномерность

распределения представленности материалов разных типов и тематик в корпусе. Обладание корпуса такими свойствами как репрезентативность и сбалансированность создает разнообразный текстовый корпус [14].

Репрезентативный корпус с большим объемом данных может гарантировать их типичность и обеспечить полноту представления всего спектра языкового явления. К корпусным данным применимы методы статистического подхода и количественный анализ, так как корпусные исследования отличаются большей представительностью данных. Единожды созданный и обработанный массив данных может быть многократно исследован в различных целях и разными исследователями. В результате корпус будет являться непрерывной средой для получения новых научных данных, что крайне необходимо в современной научной деятельности [14].

Крупнейшим российским корпусом текстов является Национальный корпус русского языка³, характеризующийся сбалансированным составом текстов. Это значит, что он по возможности содержит все типы письменных и устных текстов, представленных в русском языке, и все эти тексты содержатся пропорционально их доле в языке соответствующего периода. Корпус также содержит дополнительную разметку (или аннотацию). В отличие от электронной библиотеки, национальный корпус содержит интересные и полезные тексты с точки зрения изучения языка. В настоящее время используется пять типов разметки. Богатая и разнообразная разметка корпуса повышает его научную и учебную ценность [10].

В 1990 году одним из первых реализованных проектов с применением цифровых технологий для расширения библиотечной сферы деятельности стал проект Библиотеки Конгресса Соединенных Штатов Америки под названием «Память Америки» с последующим распространением на электронных носителях в школах и библиотеках. А в конце 1999 года совместными усилиями США и России был открыт двуязычный проект

³ Национальный корпус русского языка (НКРЯ) [Электронный ресурс] – Режим доступа: <http://ruscorpora.ru>

«Встреча на границах»⁴. Этот проект был представлен веб-ресурсом, на котором размещались исторические материалы заселения западного побережья США и освоения востока России, таким образом количество размещенного контента превысило 2,5 тыс. письменных источников, около 70 тыс. изображений из редких книг, рукописей, фотографий, карт и аудиовизуального контента, принадлежащих Библиотеке Конгресса. В последующие годы партнерами этого проекта стали более 20 библиотек, архивов и музеев городов Дальнего Востока и Сибири. Главным образом проект направлен на сохранение исторических материалов об освоении американского Запада, Сибири и Дальнего Востока, также представлены материалы о встрече русских и американцев на Аляске. В эту цифровую библиотеку вошли следующие типы источников данных: книги и периодические издания, рукописи, фотоматериалы, графика, ноты, фильмы и звукозаписи, выставки [3].

1.2 Современные электронные библиотеки

Лев Манович говорит о том, что современные художественные музеи создали строгую систему деления визуального культурного наследия на две категории: искусство и всё остальное. Под искусством понимается общепризнанные канонизированные работы преимущественно европейских мастеров. Всё, попадающее в категорию искусство, упорядочивается по именам художников, периодам времени, месту создания и стилю работы. Поэтому сегодня цифровые коллекции художественных музеев выглядят упорядоченно и систематизировано с четкой классификацией объектов искусства. В свою очередь это привело к некоторому стандарту, которому учреждения культуры должны соответствовать [1].

В случае с цифровыми библиотеками и агрегаторами появляется возможность поиска артефактов и предметов древности наравне с

⁴ «Meeting of Frontiers»

современным визуальным искусством. Вместо небольших выставок обычно посвященных одному жанру или одному наиболее влиятельному художнику в физических музеях, в цифровых коллекциях получение доступа к тысячам художественных объектов и предметов истории осуществляется за считанные секунды. Просматривая на одном веб-ресурсе нескончаемое количество объектов оцифрованных коллекций, которые были перенесены из библиотек, может создаваться впечатление о неперiodичном и несистематизированном прошлом. Бесконечно огромное наследие человеческих материальных культур было собрано и сохранено в многочисленных библиотеках, и не всегда эти объекты могут соответствовать четким критериям классификационной системы. В настоящее время они оцифрованы, а затем приведены к общим стандартам формирования метаданных, веб-протоколов и стали связаны кодом Javascript и API [1].

Реализация многочисленных проектов, схожих с Европейской, способствует созданию крупномасштабных международных программ и проектов, например, проект «Цифровое культурное наследие»⁵, организованный для сохранения всемирного массива цифровых ресурсов с предоставлением доступа к нему. В данный проект входят как оцифрованные объекты, так и изначально электронные. В результате создания такого рода проекта компанией Google в 2011 году был запущен портал Google Art & Culture⁶. Еще одним примером международного хранилища цифровых данных может выступать проект Всемирной электронной библиотеки⁷, который был создан эгидой ЮНЕСКО с участием 40 государств-членов организации, крупнейших мировых компаний, таких как Google и Microsoft. Запуск Всемирной электронной библиотеки осуществился в 2009 году [15].

На сегодняшний момент мало исследовательских работ освещающих проблемы формирования кластеров контента, размещаемого в электронных

⁵ Digital Cultural Heritage

⁶ Google Arts & Cultures [Электронный ресурс] – Режим доступа: <https://artsandculture.google.com/>

⁷ World Digital Library

библиотечных системах, а также исследований существующих технологий из других областей науки, которые способны справиться с выявленными проблемами.

В своей книге «Cultural Analytics» Лев Манович рассуждает о больших объемах культурных данных не как о недостижимой роскоши, а скорее как о вынужденной необходимости, которая должна обязательно присутствовать, если мы хотим изучать культурное разнообразие в зависимости от географии, времени, авторства, средств коммуникации и других характеристик [16].

1.3 Цифровая библиотека Европеана

Европеана - европейская цифровая библиотека, в приоритетные задачи которой входит сохранение и предоставление открытого доступа к информации об объектах, представляющих ценность в качестве сохранения и актуализации культурной памяти [17].

В июне 2005 года Европейская Комиссия приняла решение о создании европейской цифровой библиотеки для роста информационного общества как стратегию ЕС на 2010 год [18]. Проект Европеана (оригинальное название - Europeana) был запущен 20 ноября 2008 года, на первых этапах веб-ресурс включал информацию о примерно 4,5 млн объектов культурного наследия в оцифрованном виде. Главный сервер расположен в Нидерландах и его работу контролирует Национальная библиотека Нидерландов⁸.

При формировании цифровых коллекций Европеаны были включены документы в текстовом формате, фотографии, мультимедийный контент о предметах музейного фонда, материалы архивов, звуковые и аудиовизуальные файлы, электронные версии археологических объектов и артефактов [17].

Запуск такого крупного европейского агрегатора способствует укреплению связей в мультиэтничном населении Европы, а также создает

⁸ Koninklijke Bibliotheek [Электронный ресурс] – Режим доступа: <https://www.kb.nl/>

взаимовыгодные партнерские отношения между различными учреждениями культуры.

Европеана дает возможность рядовым интернет-пользователям, не обладающим профессиональным образованием/специальной подготовкой, получать доступ к огромному массиву культурно и исторически важных данных, с целью использования их в учебных, исследовательских, профессиональных и развлекательных целях. Также многие размещаемые документы имеют открытые лицензии, что упрощает работу для исследователей. Особенностью является то, что она лишь предоставляет доступ и создает общий интерфейс для всех объектов, а не хранит их. Все данных хранятся в музеях, архивах и других учреждениях культуры, которые и предоставили свои объекты культурного наследия.

По данным на ноябрь 2021 в библиотеке Европеана опубликованы тематические коллекции, содержащие в сумме более 62 миллионов оцифрованных объектов. Также включены 14 693 русскоязычных документа⁹ (14 665 текстового типа), 25 из которых предоставлены Украиной. Поставщиками русскоязычного контента являются: Русская государственная библиотека, Архангельская областная научная медицинская библиотека, Национальная библиотека Удмуртской Республики, Государственный музей Л.Н. Толстого и др.

Данный проект имеет огромное значение в контексте реализации задач по сохранению мирового культурного наследия и предоставлению глобального доступа к мировым информационным ресурсам. Также значительный рост и расширение аналогичных проектов позволяют сделать вывод о том, что такой способ организации доступа к информационным ресурсам, преимущества в библиотечной среде и в учреждениях культуры разных мировых стран получил большое распространение [15].

⁹ данные на 20 марта 2022 года

Может ли европейская цифровая библиотека являться репрезентативной и сбалансированной основой и содержать разнообразный контент для гуманитарных исследований?

Создатели и разработчики Европейцы сами не занимаются оцифровкой материалов для коллекций. Вместо этого они лишь предоставляют доступ к большим коллекциям данных, которые им предоставили музеи и другие учреждения культуры. В результате такого взаимодействия цифровой корпус данных отражает то, что собирают учреждения культуры и что, по их мнению, важно и должно быть оцифровано в первую очередь. Доподлинно неизвестно какими принципами руководствуются редакторы при отборе объектов для публикации. Таким образом целью цифровых коллекций, в частности Европейцы, обычно бывает распространение знаний о культурах мира, а достижение этой цели не гарантирует репрезентативность, сбалансированность и разнообразие собранного корпуса.

2 Материалы и методы исследования

2.1 Методы исследования

По мере развития таких цифровых библиотек как Europeana, Google Arts & Culture, Digital Public Library of America (DPLA) и других, фундаментальный вопрос о количественном подходе к исследованию материалов истории культуры различных этнических групп со всего мира остается без ответа. Главным образом нерешенной является задача формирования репрезентативной выборки, которая, в свою очередь, будет систематизировать то, что было создано в одном временном периоде, в одной географической области, стиле, жанре и других категориях. Методы и подходы к систематизации и точной классификации данных четко сформировались в среде компьютерных и социальных наук, но не в гуманитарных областях. Поскольку учебники, фильмы, фотографии, учреждения культуры, а главное - школьные и институтские лекционные занятия, представляют человеческое искусство и культуру только на анализе и группировке отобранных вручную работ, идея о создании систематизированной и репрезентативной выборки становится, как никогда, злободневной [1].

Проведенные неоднократно эксперименты доказывают, что качественный метод исследования наполнения цифровых коллекций будет сопровождаться предубеждением в отношении социальной желательности. Это значит, что при опросе респондентов будут прослеживаться тенденции представления социального контекста таким образом, который будет

восприниматься как социально приемлемый в этом обществе [19]. Это может привести к переоценке полученных данных в более позитивную сторону и уменьшению неоднородности ответов респондентов, что приведет к необъективной оценке собранных данных. В нашем случае, опираясь на данные, собранные качественным методом, мы не сможем дать четко сформулированную оценку, так как визуальный интерфейс библиотеки Европейца демонстрирует сбалансированность и разнообразие тематических коллекций. Поэтому для получения объективных данных по тематическим коллекциям Европейца будет использован количественный метод анализа.

Несмотря на то что, цифровые коллекции обычно стремятся распространять знания о культурах мира, степень репрезентативности, сбалансированности и разнообразия таких выборок данных остается по-прежнему неизвестной. В текущей работе будет предпринята попытка оценить сбалансированность географического распределения материалов коллекций и представленность 197 стран мира количественным методом анализа данных. Все ли страны представлены одинаково в проекте европейской цифровой библиотеке? Какие страны и учреждения культуры являются лидерами по предоставлению доступа к оцифрованным материалам естественно-научного и культурного наследия?

Разнообразие контента будет оцениваться представленностью различных тематик внутри крупных разделов в онлайн базе данных Европейца. Для этого будет проведен количественный анализ представленности выделенных тематик.

Формирование представлений об иной культуре, которое будет основываться на недостаточном количестве данных, искаженных данных или данных, которые представляют культурное наследие сквозь призму западного восприятия, неизбежно приводит к замедлению взаимного обогащения культур и социальной мобильности, а также культурный капитал перестает расширяться. Культурный анализ, построенный на такого рода больших данных, способствует искажению и неверной трактовке

получаемых результатов. Таким образом наиболее демонстрируемые тенденции и закономерности характерные для культур, представленных в агрегаторах и цифровых коллекциях, будут вытеснять особенности представленных в недостаточном количестве культур [20].

Цифровая электронная библиотека Европеана обладает обширным количеством оцифрованных данных различных видов (текст, изображение, видео, звук, 3D модель). Вследствие постоянного расширения базы данных новыми электронными документами и количественным увеличением объемов 13 основных коллекций Европеаны, библиотека сталкивается с проблемой оснащения сервиса эффективным поиском по массивам данных в условиях сложной архитектуры метаданных об объектах культурного наследия, в свою очередь Европеана успешно решает эту проблему давая пользователям возможность фильтрации данных по многим критериям. Применение метода фасетного поиска для разноаспектной фильтрации служит удобным механизмом построения поисковых запросов, а также позволяет сокращать время, затраченное на поиск необходимых данных определенного типа.

2.2 Сбор и обработка данных

Разработчики сайта Европеаны 20 декабря 2021 года обновили старую версию веб-ресурса. Главным образом произошли изменения в облике главной страницы, а также изменился интерфейс и стали доступны новые возможности перемещения и сортировки в коллекциях.

В текущем исследовании анализ данных проводился с использованием двух версий Европеаны для того, чтобы в полной мере рассмотреть затронутые проблемы.

Начальные страницы старой и новой версии представлены на рисунке 1 и 2.

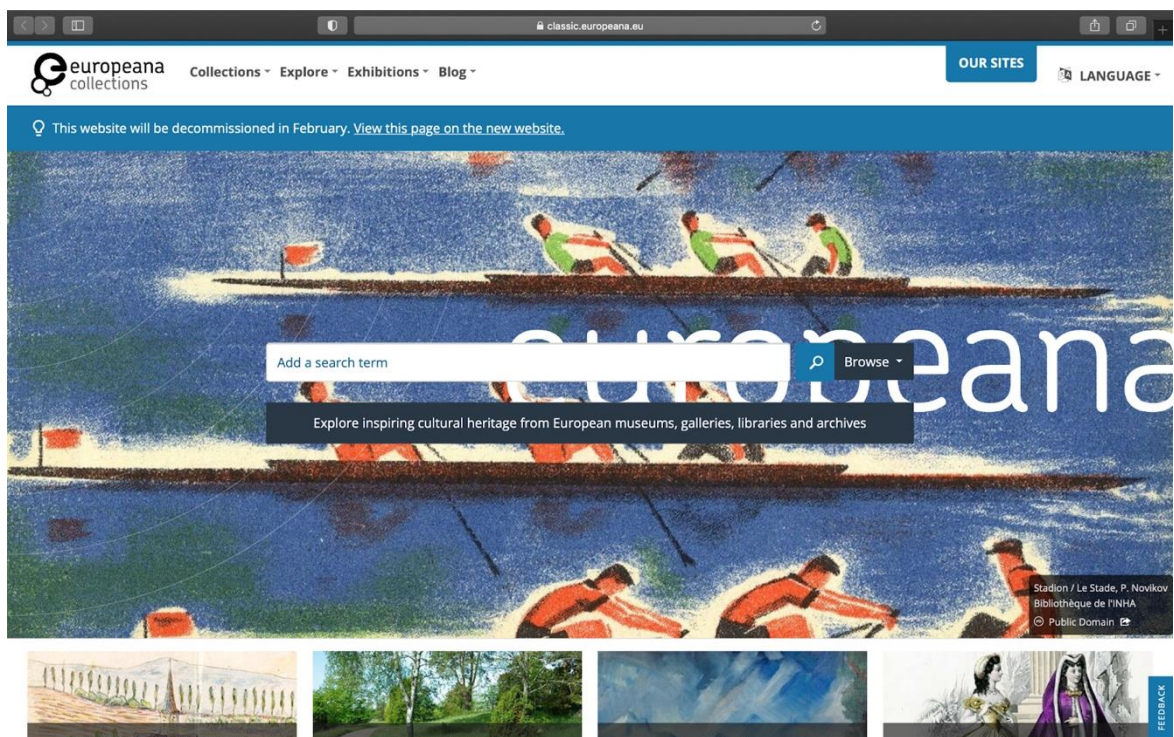


Рисунок 1 - Главная страница старой версии сайта Европеаны

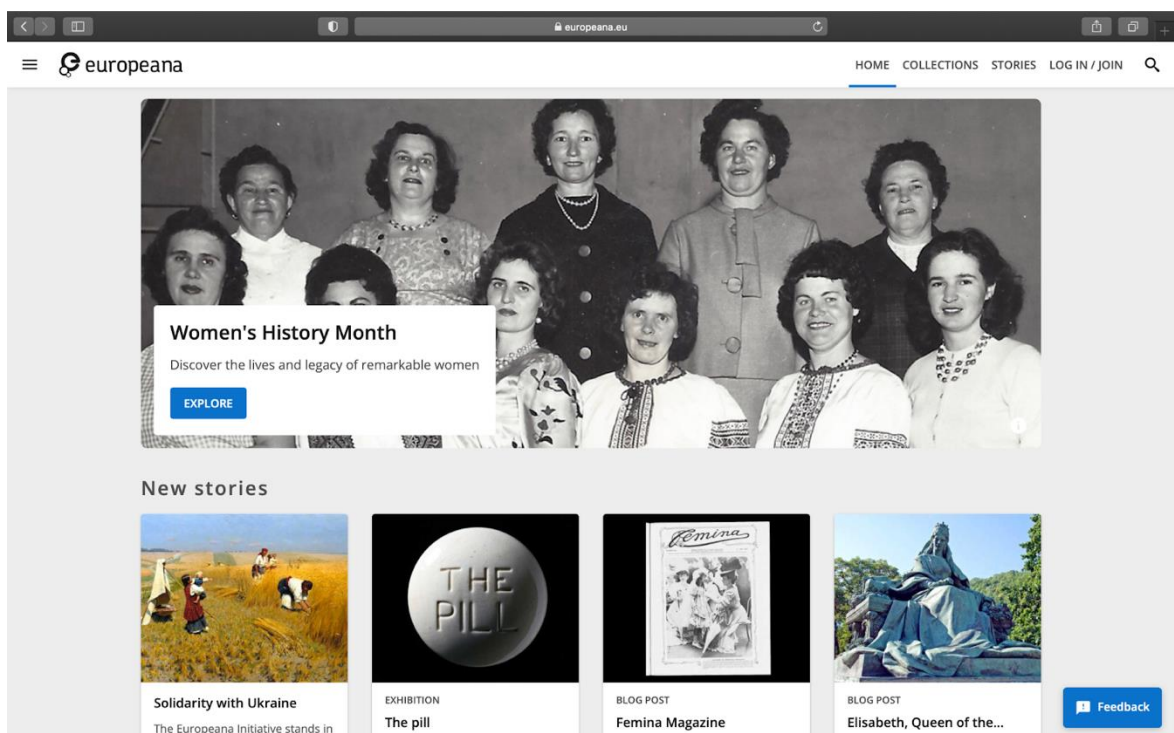


Рисунок 2 - Главная страница новой версии сайта Европеаны

Интерфейс обеих версий позволяет получать быстрый доступ к 13 основным коллекциям, созданных специально для проекта Европеаны.

Основные коллекции и количество материалов в каждой из них представлено на рисунке 3.

1914-1918	380 933
Archaeology	2 612 149
Art	3 143 755
Fashion	904 589
Industrial Heritage	801 216
Manuscripts	547 489
Maps and Geography	84 161
Migration	301 226
Music	906230
Natural History	10 352 422
Newspapers	995 182
Photography	5 882 984
Sport	306 674

Данные от 01.11.2021

Рисунок 3 - 13 основных тематических коллекций Европеаны и количество учетных записей в каждой коллекции. Автор: Арина Орлова

Интерфейс старой версии библиотеки позволяет получать пользователям доступ не только к 13 основным коллекциям, но и к другим разделам, которые содержат:

- 173 тематические подборки по разным типам материалов;
- 83 статьи, посвященные известным личностям разных эпох;
- материалы по 13 временным периодам (начиная с 13 века и до нашего времени);
- 47 тематических выставок;
- 672 поста в блоге;
- 215 тематических фотоподборок.

На ноябрь 2021 года совокупный объем онлайн базы данных электронной библиотеки Европеана составил 62 872 673 учетные записи. В свою очередь Европеана включает в себя материалы как крупных библиотек,

музеев и галерей, так и совсем небольших коллекций из городских или личных архивов. В данном исследовании проводится анализ онлайн базы данных европейского цифрового агрегатора Европеаны, в котором хранятся материалы цифровых коллекций со всего мира (см. рис. 4).

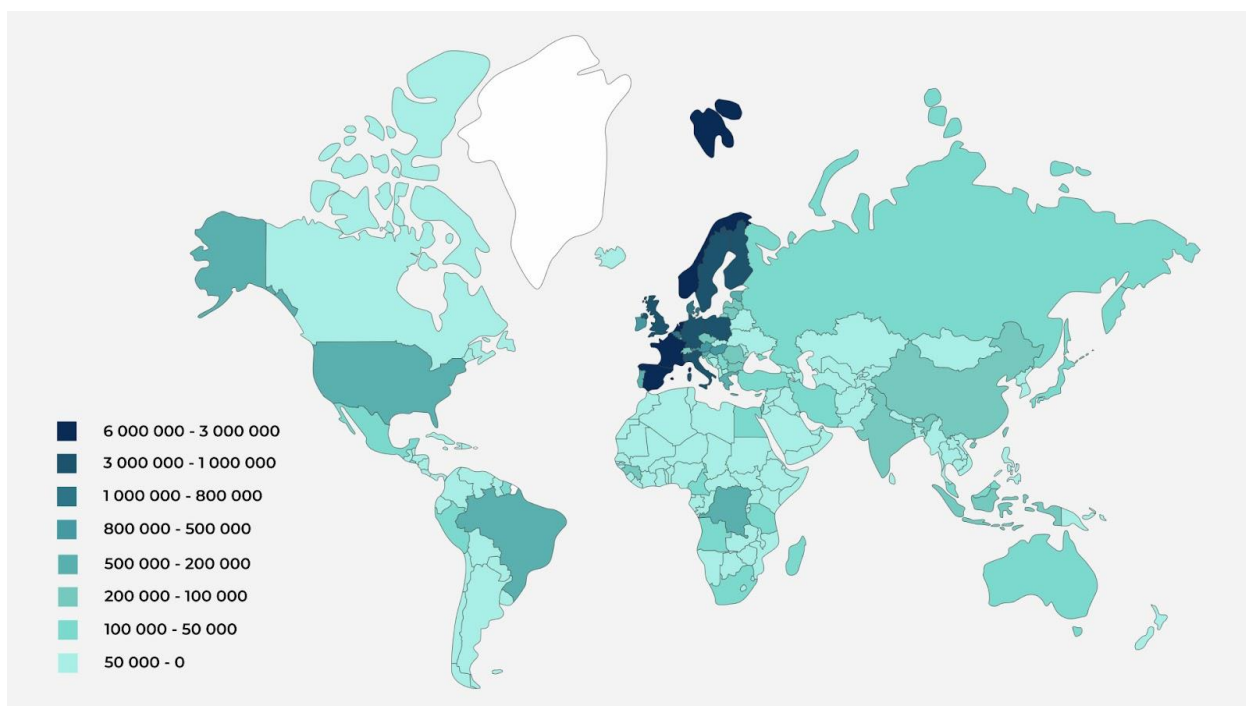


Рисунок 4 - Тепловая карта мира, отражающая распределение оцифрованных материалов, хранящихся в онлайн базе данных Европеаны. Автор: Арина Орлова

2.2.1 Разнообразие контента

Для оценки разнообразия представленных в Европеане тематических коллекций в ноябре 2021 года необходимо было определить количество экспонатов в 13 крупнейших коллекциях. На основании собранных данных, появляется возможность вычисления процентного соотношения каждой коллекции по отношению к общему количеству учетных записей, объем которых составляет 62 872 673 учетные записи.

Более детальный разбор каждой коллекции включал в себя выделение 5 лидеров среди организаций и стран, которые открыли доступ к свободному

бесплатному просмотру и использованию оцифрованного культурно-исторического контента в каждом крупном тематическом разделе Европеаны.

Данные об организациях дают возможность определить представленные коллекциями страны и учреждения культуры, ведь специализация того или иного учреждения напрямую влияет на хранящиеся в них культурно-исторические материалы и определяет их вид. В работе отражено, какие страны чаще всего предоставляют свои материалы для агрегации и как представлены все страны из списка стран-членов ООН по количеству учетных записей - все ли страны представлены одинаково или имеется скос в сторону определенной тематической направленности размещаемых Европейой материалов и конкретного типа материалов. Подсчет количества учетных записей и выделение 5 учреждений культуры в основных крупных коллекциях поможет определить как распределены материалы внутри коллекций, а также выявить лидеров среди поставщиков цифрового естественно-научного и культурного наследия в онлайн базе данных Европеаны.

Полученные данные о 5 учреждениях культуры, предоставивших свои оцифрованные материалы в каждой из 13 коллекций, внесены в отдельную Excel таблицу. Часть таблицы Excel представлена на рисунке 5. В таблице представлены следующие данные:

1. Название коллекции;
2. Количество объектов в коллекции;
3. Названия 5 музеев;
4. Количество объектов в коллекции, предоставленное каждым музеем;
5. Процентное соотношение количества объектов одного музея по отношению к общему количеству материалов коллекции.

Разделы	Кол-во записей в коллекции	Музей 1	Кол-во объектов в музее 1	Процент объектов в музее	Музей 2	Кол-во объектов в музее 2	Процент объектов в музее	Музей 3	Кол-во объектов в музее 3	Процент объектов в музее	Музей 4	Кол-во объектов в музее 4	Процент объектов в музее	Музей 5	Кол-во объектов в музее 5	Процент объектов в музее
1914-1918	369 896	Österreichische Nationalbibliothek - Austrian National Library	97 892	26,5%	Central Institute for the Union Catalogue of Italian Libraries	56 137	15,2%	National Library of France	41 852	11,3%	Europeana 1914-1918	32 054	8,7%	National Library of Spain	30 640	8,3%
Archaeology	2 549 924	Swedish National Heritage Board	560 880	22,0%	Beeldbank van de Rijksdienst voor het Cultureel Erfgoed	500 718	19,6%	The National Heritage Institute	251 542	9,9%	The Danish Agency for Culture	131 405	5,2%	Deutsches Archäologisches Institut	104 659	4,1%
Art	3 044 096	Rijksmuseum	337 939	11,1%	German Documentation Center for Art History - Marburg Picture Index	210 770	6,9%	Österreichische Nationalbibliothek - Austrian National Library	199 341	6,5%	National Library of France	114 960	3,8%	Albertina	97 431	3,2%
Fashion	905 149	Catwalkpictures	514 919	56,9%	Nordic Museum Foundation	56 469	6,2%	Paul Van Riel	45 617	5,0%	Fashion Museum of Antwerp	37 952	4,2%	National Museum of World Cultures Foundation	28 482	3,1%
Industrial Heritage	830 935	Swedish National Museum of Science and Technology	56 453	6,8%	Central Institute for the Union Catalogue of Italian Libraries	51 822	6,2%	The National Archives of Norway	51 142	6,2%	Österreichische Nationalbibliothek - Austrian National Library	50 018	6,0%	Deutsche Fotothek	37 531	4,5%
Manuscripts	552 679	National Library of France	88 859	16,1%	National Library of the Czech Republic	76 032	13,8%	Bodleian Libraries, University of Oxford	72 525	13,1%	National Library of the Netherlands	38 717	7,0%	The British Library	33 604	6,1%

Рисунок 5 - Таблица с информацией о количестве учетных записей в 13 основных коллекциях Европеаны, а также процентное распределение 5 учреждений культуры, представивших максимальное количество материалов в каждой из коллекций, по отношению к общему количеству материалов в каждой из 13 коллекций. Автор: Арина Орлова

В целях установления степени репрезентативности тематических выборок европейской цифровой библиотеки Европеаны, следовало провести анализ и оценить разнообразие тематик контента внутри крупных разделах агрегатора, которые расположены на главном интерфейсе. Сбор данных о количестве материалов, статей и документов в них проводился в январе 2022 года. Разделы интерфейса с указанием количества учетных записей в каждом представлены на рисунке 6. Последующий количественный анализ наполнения тематических разделов проводился методом случайного отбора 20% учетных записей или использовались все учетные записи в конкретном разделе, если он содержал небольшое количество материалов. Далее отобранным учетным записям присваивалась достаточно широкая тематическая категория в зависимости от наполнения этой записи. Анализу были подвергнуты следующие тематические разделы: “Topics”, “Galleries”, “Exhibitions”, “Blog Posts”.

Collections	Explore	Exhibitions	Blog
1914 - 1918 (369 889)	What's new?	Exhibitions (47)	Blog posts (672)
Archaeology (2 540 368)	Colours (138 colours)		
Art (2 995 577)	Sources (220 institutions)		
Fashion (905 149)	Topics (173)		
Industrial Heritage (787 011)	People (93)		
Manuscripts (555 812)	Time periods (12)		
Maps and Geography (56 179)	Galleries (215)		
Migration (302 994)			
Music (919 119)			
Natural History (10 419 817)			
Newspapers (4 140 220)			
Photography (6 465 228)			
Sport (314 460)			

Данные за 17.01.2022 - 31.01.2022

Рисунок 6 - Главные разделы старой версии Европеаны с указанием количества учетных записей в каждом из разделов. Автор: Арина Орлова

В старой версии Европеаны разработчиками вводилось внутреннее тегирование больших разделов на небольшие подтемы. Эти темы представлены на рисунке 7.



Рисунок 7 - Тематические категории, на которые Европейана разбивает более крупные коллекции. Автор: Арина Орлова

Отдельно были проанализированы 113 записей, размещенных в разделе тематических подборок “Topics”, что составило 65,3% от общего объема тематических подборок в этом разделе. Была составлена отдельная таблица со следующими данными (см. рис.8):

- Название тематической подборки;
- Количество записей в подборке;

- Обобщенная тематика записей, представленных в подборке;
- Доминирующий тип материалов;
- Комментарии с развернутым указанием материалов в подборке.

№	Название	Кол-во записей	Тематика	Типы материалов	Комментарии
1	Aerial photography	350	Photography	Photography	Фотографии с воздуха
2	album leaf	249	Photography	Photography	Черно-белые фотографии и негативы из театра, альбомные фотографии из семейной жизни, меню из отелей
3	Alternative medicine	168	Medicine	Books, photography	Ботанические фотографии и иллюстрации, зарисовки и фотографии лечения пациентов, книги о народной медицине
4	Ambrotypes	73	Photography	Photography	Черно-белые фотографии из личных архивов
5	Animal husbandry	5 454	Animals, tools	Photography	Фотография и графика на тему животноводства, применяемые инструменты
6	Archaeological site	460	Tomb (archaeology)	Photography	Рабочие материалы с мест раскопок
7	Art	2 065 868	Art (Paintings, drawings, engravings and sculpture from cultural heritage institutions across Europe.)	Photography, text, sound	Большой раздел со всем европейским искусством. Много фотографий колониального прошлого
8	Artifact (archaeology)	111	Artifact (archaeology)	Photography	Археологические находки разных периодов
9	Avant-garde music	3	Music	Text	Статьи
10	Baroque painting	3	Art (history)	Text, sound	Книги о искусстве периода Барокко
11	Biography	10 783	Biography	Books	Книги и статьи
12	Black-and-white photographs	24 070	Photography	Photography, text	Городские фотографии, открытки
13	Blues	2 668	Music	Sound	Музыка из интернет архивов
14	Boogie rock	185	Music	Sound	Музыка

Рисунок 8 - Скриншот Excel таблицы с информацией о 113 тематических подборках в разделе “Topics”. Автор: Арина Орлова

В обновленной версии европейской цифровой библиотеки Европеаны с 20 декабря 2021 года появилась возможность обращения и просмотра материалов онлайн базы данных по вековым периодам начиная с I века. Используя эту возможность, была составлена Excel таблица представляющая информацию о количестве разных типов материалов по всем векам, также по каждому веку отобраны 5 учреждений культуры и 5 стран, которые предоставили наибольшее количество материалов по определенному веку. Часть этой таблице представлена на рисунке 9.

Век	Кол-во объектов в разделе	Топ 5 музеев	Кол-во объектов уч. куль.	Топ 5 стран	Кол-во объектов по странам
1	11 090	Cyfrowy Dolny Śląsk	6 821	Poland	7 536
		National Library of Denmark	975	Denmark	1 334
		National Library of France	716	France	721
		The Royal Library: The National Library of Denmark and Copenhagen University Library	359	Portugal	339
		G.T. Kirileanu Neamt County Public Library	241	Romania	317
2	1 947	Universitat Autònoma de Barcelona	499	Spain	516
		The Royal Library: The National Library of Denmark and Copenhagen University Library	411	Denmark	432
		German Documentation Center for Art History - Marburg Picture Index	145	Poland	167
		INP – National Heritage Institute, Bucharest	106	Romania	156
		National Library of France	93	Germany	146
3	462	German Documentation Center for Art History - Marburg Picture Index	142	Romania	150
		INP – National Heritage Institute, Bucharest	96	Germany	143
		Eesti Sõjamuuseum - Kindral Laidoneri Muuseum	59	Estonia	59
		National Museum of Romanian History	37	Netherlands	22
		Archaeology Data Service	14	United Kingdom	17

Рисунок 9 - Скриншот Excel таблицы с рейтингом 5 учреждений культуры и стран, отсортированных по векам, с информацией о количестве учетных записей по ним. Автор:

Арина Орлова

2.2.2 Сбалансированность данных

Для оценки сбалансированности географической представленности стран мира внутри онлайн базы данных Европеаны составлен список из 197 стран, включающий страны-члены ООН. Сбор данных о количестве учетных записей проводился с помощью встроенного в веб-ресурс поиска без применения встроенной фильтрации. Подсчет общего числа экспонатов, посвященных различным странам мира, позволит оценить представленность каждой страны в Европеане. Эти данные позволят выявить лидеров оцифрованного контента, а также страны, материалы по которым отсутствуют. Все собранные данные были оформлены в Excel таблицу, которая содержит информацию о количестве учетных записей по каждой стране из списка (см. рис. 10).

Страна	Кол-во учетных записей
USA	263 484
Italy	1 497 992
Great Britain	544 166
United Kingdom	1 051 472
Spain	3 211 861
Germany	1 069 558
France	5 150 283
Japan	78 314
Brazil	325 714
India	168 451
Korea	11 581
Russia	97 450
Australia	94 785

Рисунок 10 - Скриншот Excel таблицы с информацией о количестве учетных записей в онлайн базе данных Европеаны. Автор: Арина Орлова

При вводе в поисковой строке названия страны, конечный результат будет отображать все учетные записи, в метаданных которых есть упоминания этой страны. Учреждения культуры какой-либо страны могли отказаться работать и предоставлять свои объекты наследия для размещения их в Европеане. Но другие страны в ходе культурно-исторического взаимодействия друг с другом могли получить в свое владение материалы других стран. Например, в эпоху колониальных и постколониальных культур в европейских архивах и библиотеках накапливалось огромное количество гербариев, которые были собраны на новых завоеванных землях. В этом случае такие объекты естественно-научного и культурного наследия могут быть тегированы страной происхождения. Таким образом получится рассмотреть представленность стран и культур в цифровой библиотеке.

Дальнейшая работа заключалась в установлении скопа в учреждениях культуры, которые отправляют материалы своих коллекций для размещения в самой крупной цифровой коллекции Европеаны - "Natural History". Для

того чтобы появилась возможность оценить скос и составить представление о лидерах стран и учреждениях агрегации различных типов документов в категории естественной истории, была собрана информация с помощью поисковых запросов с названиями стран в коллекции “Natural History”. Пример применение поиска по коллекции размещен на рисунке 11.

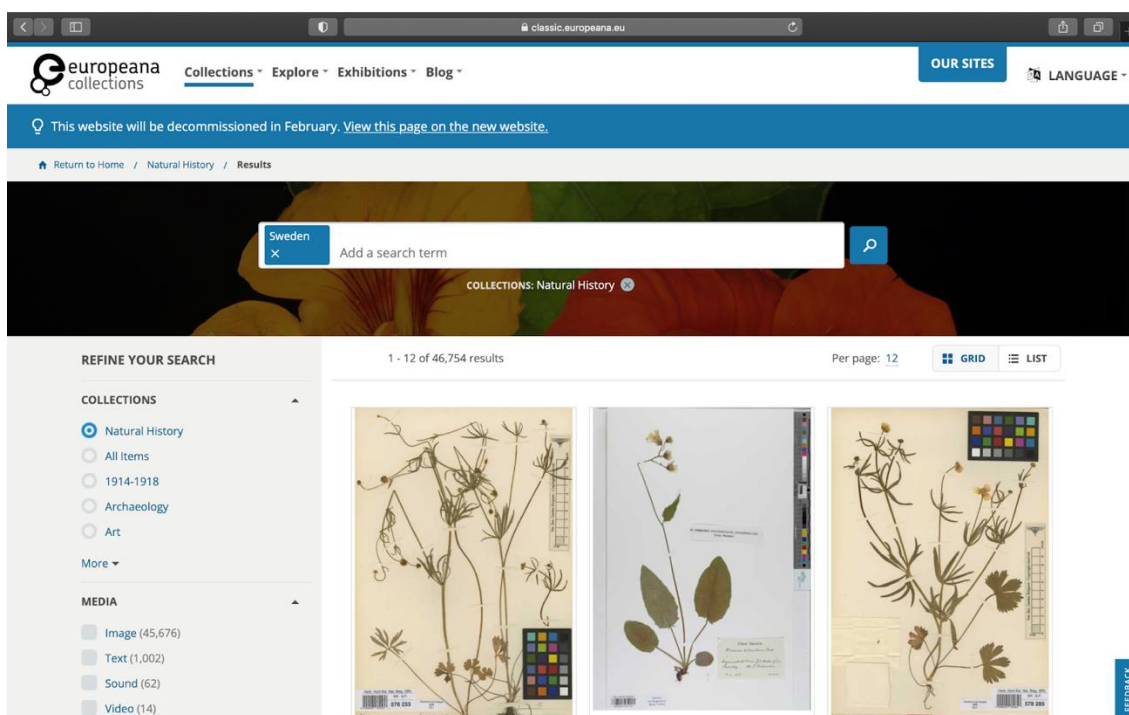


Рисунок 11 - Скриншот применения фасетного поиска для получения информации о количестве учетных записей в каждой из 197 стран ООН. Автор: Арина Орлова

После ввода необходимого запроса в поисковую строку в поле фильтров данных появляется возможность увидеть рейтинг из 5 учреждений, предоставивших наибольшее количество оцифрованных материалов, которое указывается в скобках, по введенному запросу (см. рис. 12). При открытии выпадающего списка видны все учреждения культуры, материалы которых были тегированны соответствующей страной.

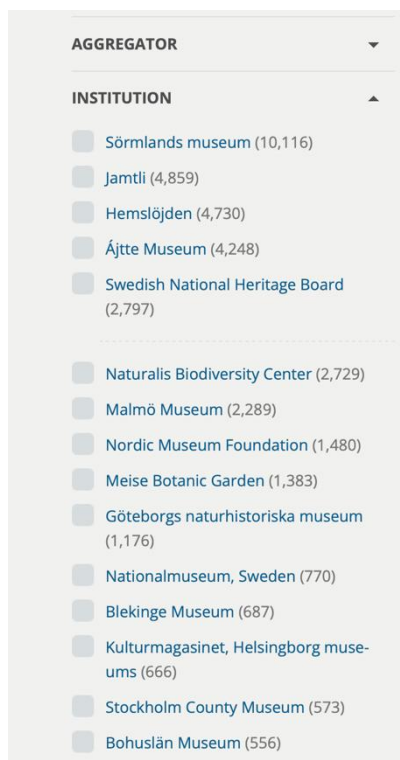


Рисунок 12 - Список учреждений культуры в категории “Natural History” при вводе в поисковой строке страны “Sweden”. Автор: Арина Орлова

Полученная в ходе поиска информация по географическому расположению материалов естественнонаучной коллекции “Natural History” была внесена в Excel таблицу, содержащую информацию о количестве учетных записей по каждой из стран и список из пяти учреждений, которые разместили наибольший объем материалов (см. рис. 13). Данные были собраны в ноябре 2021 года.

USA	38 551	Royal Botanic Gardens, Kew	Naturalis Biodiversity Centre	U.S. Department of Agriculture, National Agricultural Library	Muséum national d'Histoire naturelle	Meise Botanic Garden
Italy	107 690	Department of Life Sciences, University of Trieste	Naturalis Biodiversity Centre	The Trustees of the Natural History Museum (London)	Meise Botanic Garden	Botanic Garden and Botanical Museum Berlin
Great Britain	49 469	The Trustees of the Natural History Museum (London)	The British Library	Royal Botanic Gardens, Kew	U.S. Department of Agriculture, National Agricultural Library	Naturalis Biodiversity Centre
United Kingdom	56 825	The Trustees of the Natural History Museum (London)	The British Library	Royal Botanic Gardens, Kew	Naturalis Biodiversity Centre	U.S. Department of Agriculture, National Agricultural Library
Spain	47 124	Naturalis Biodiversity Centre	Meise Botanic Garden	Natural History Museum, Vienna Herbarium W	The Trustees of the Natural History Museum (London)	Royal Botanic Gardens, Kew
Germany	215 589	Botanic Garden and Botanical Museum Berlin	Museum fuer Naturkunde Berlin, Tierstimmenarchiv	Museum für Naturkunde Berlin	Naturalis Biodiversity Centre	Department of Life Sciences, University of Trieste
France	93 730	Naturalis Biodiversity Centre	Meise Botanic Garden	Muséum national d'Histoire naturelle	The Trustees of the Natural History Museum (London)	Department of Life Sciences, University of Trieste
Japan	19 847	Naturalis Biodiversity Centre	Royal Botanic Gardens, Kew	Muséum national d'Histoire naturelle	The Trustees of the Natural History Museum (London)	Meise Botanic Garden
Brazil	300 325	Royal Botanic Gardens, Kew	Natural History Museum, Vienna - Herbarium W	Meise Botanic Garden	Muséum national d'Histoire naturelle	Naturalis Biodiversity Centre

Рисунок 13 - Скриншот Excel таблицы с информацией о количестве учетных записей в коллекции “Natural History” по 197 странам в Европе. Автор: Арина Орлова

2.3 Визуализация данных исследования

Визуализация всех полученных в ходе исследования данных производилась в графическом онлайн-редакторе Figma. Данный инструмент заточен под веб-дизайн и имеет меньший набор функций, чем, например, Adobe Photoshop, не требует установки на ПК, а позволяет через браузер подключать других пользователей для совместной работы или демонстрации уже готового продукта. Более подробно плюсы работы в Figma разбирают в своей статье Хвостенко Т.М. и Велисар Д.С.. На рисунке 14 представлен скриншот процесса визуализации данных исследования.

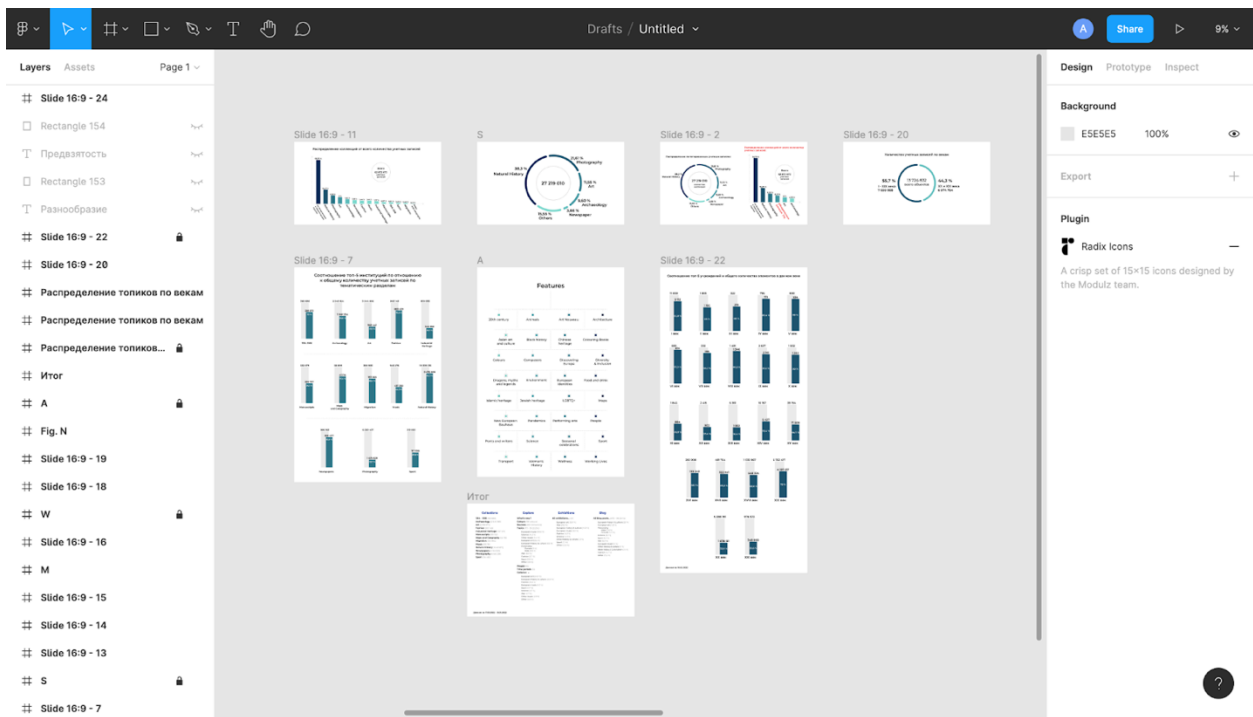


Рисунок 14 - Процесс визуализации полученных данных в онлайн графическом редакторе Figma. Автор: Арина Орлова

3 Результаты исследования

3.1 Анализ данных по разнообразию материалов коллекций Европеаны

В ходе исследования было определено, что самой крупной коллекцией в Европеане является коллекция естественной истории “Natural History”. Большинство материалов в этой коллекции представляют собой исследовательские материалы, гербарии, архивные лабораторные журналы, фотографии с мест археологических раскопок. Как было установлено, эта коллекция составляет одну треть часть от всего количества контента в Европеане и насчитывает 10 352 422 объектов в старой версии сайта (см. рис. 15).

Распределение коллекций от всего количества учетных записей

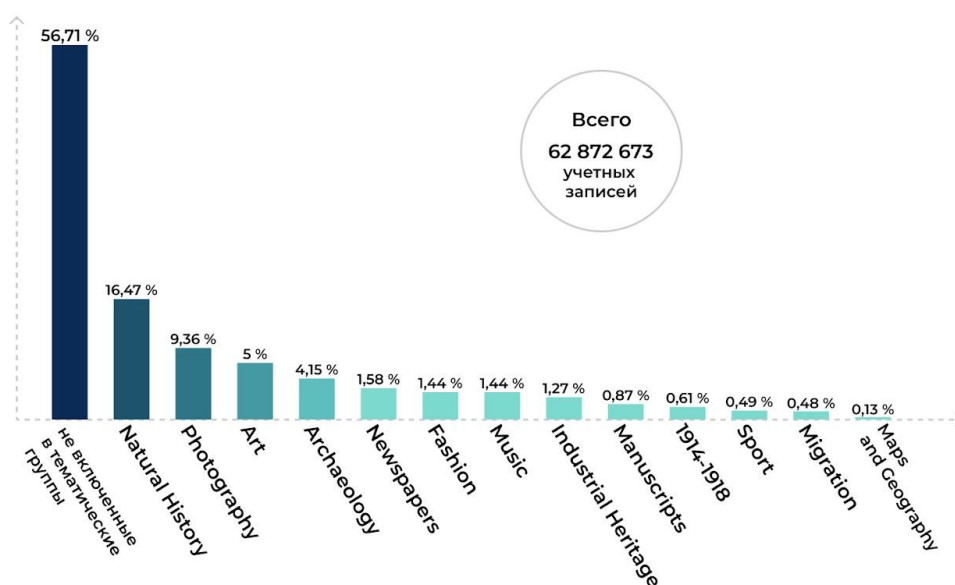


Рисунок 15 - Процентное соотношение материалов основных 13 коллекций Европеаны к общему количеству учетных записей. Автор: Арина Орлова

Общий объем 13 коллекций составил 27 219 010 объектов. На рисунке 16 визуализированы данные процентного соотношения 5 крупнейших коллекций Европеаны к общему количеству материалов по всем 13

коллекциям. В категорию “Other” вошли остальные 8 коллекций, суммарный процент по которым составил 15,55 %, а процент по каждой не превысил 3,5 %. Минимальный процент представленности материалов коллекции “Maps and Geography” составил 0,31%.

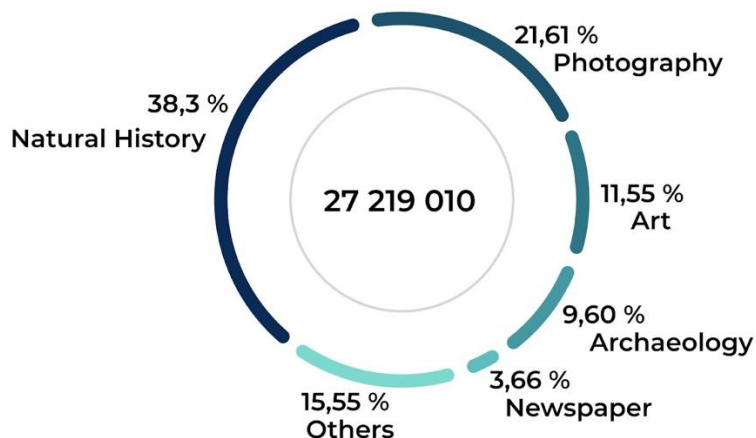


Рисунок 16 - Диаграмма процентного соотношения 13 крупных коллекций Европейцы к общему количеству тегированных материалов по всем коллекциям. Автор: Арина Орлова

Каждая тематическая коллекция включает в себя большие объемы данных из разных учреждений культуры и стран. Выделение 5 организаций по количеству размещенных материалов в каждой из коллекций по отношению к общему объему коллекции поможет выявить какой процент наполнения контента составляют материалы из первых 5 учреждений культуры по количеству предоставленных для агрегации цифровых объектов культуры и истории. Это важно, потому что больший процент распределения контента из 5 учреждений культуры по отношению к общему количеству материалов в данной категории наглядно демонстрирует снос в географическом распределении. Такого рода снос в данных способствует однородности контента как с точки зрения культурного, так и тематического разнообразия. Количественное и процентное соотношение собранных данных представлено на рисунке 17. Серые столбцы и их численные значения

отражают количество объектов различных видов данных в каждом разделе, синие же столбцы визуально демонстрируют долю контента в коллекции 5 учреждений.

Данная диаграмма позволяет нам сделать вывод о том, что в разделе “Natural History” 82,3% материалов были предоставлены для агрегации лишь 5 учреждениями культуры что значительно влияет на репрезентативность такой выборки. Это показывает малое типовое и видовое разнообразие внутри крупнейшей коллекции Европеаны.

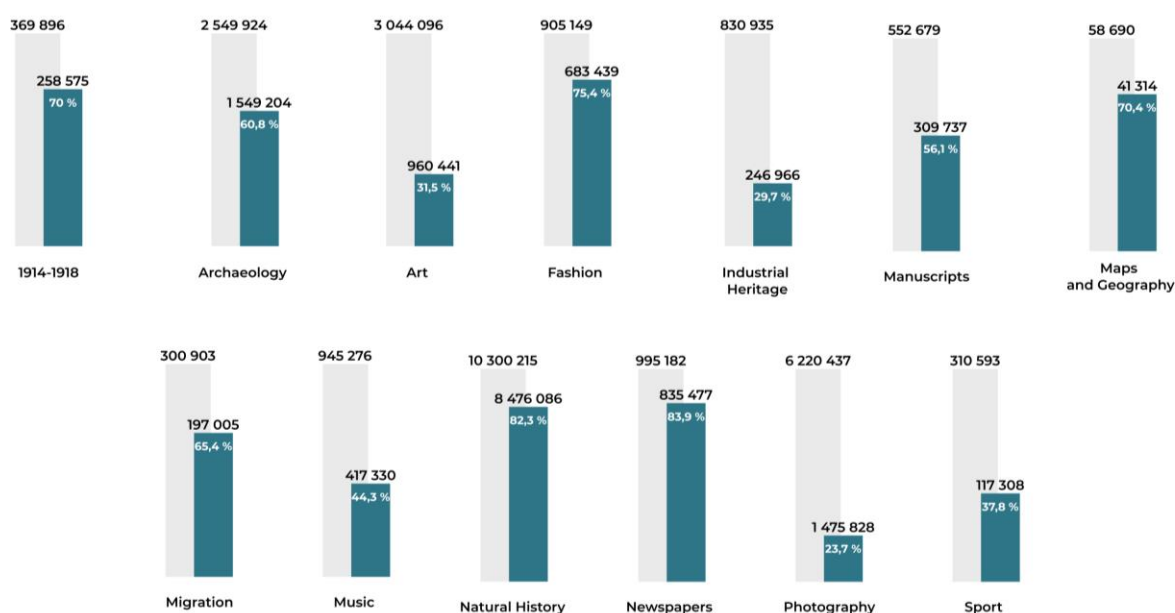
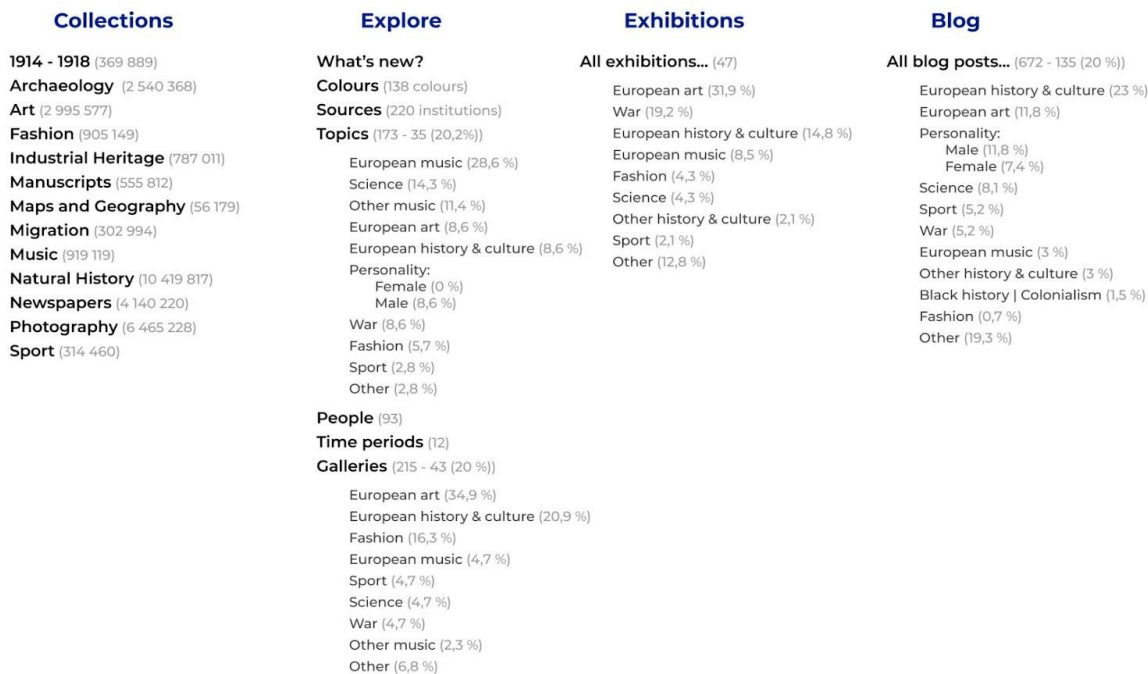


Рисунок 17 - Соотношение 5 крупных учреждений культуры по отношению к общему количеству учетных записей в 13 основных коллекциях Европеаны. Автор: Арина Орлова

Последующий анализ 4 главных разделов, размещенных в интерфейсе, проводился по сортировке отобранных 20% учетных записей в каждом разделе по обобщенным тематикам.

Результаты проведенного анализа показывают, что, хотя содержимое базы данных Европеаны можно разбить на различные тематические категории, контент не представляется однородным по количеству объектов в

каждой теме, и наша сортировка материалов по тематическим направлениям это отображает (см. рис. 18).



Данные за 17.01.2022 - 31.01.2022

Рисунок 18 - Анализ содержания главных разделов интерфейса Европеаны по тематическим категориям. Автор: Арина Орлова

Для более подробного представления результатов этого анализа была создана визуализация с указанием 5 преобладающих тем в каждом разделе (см. рис. 19).

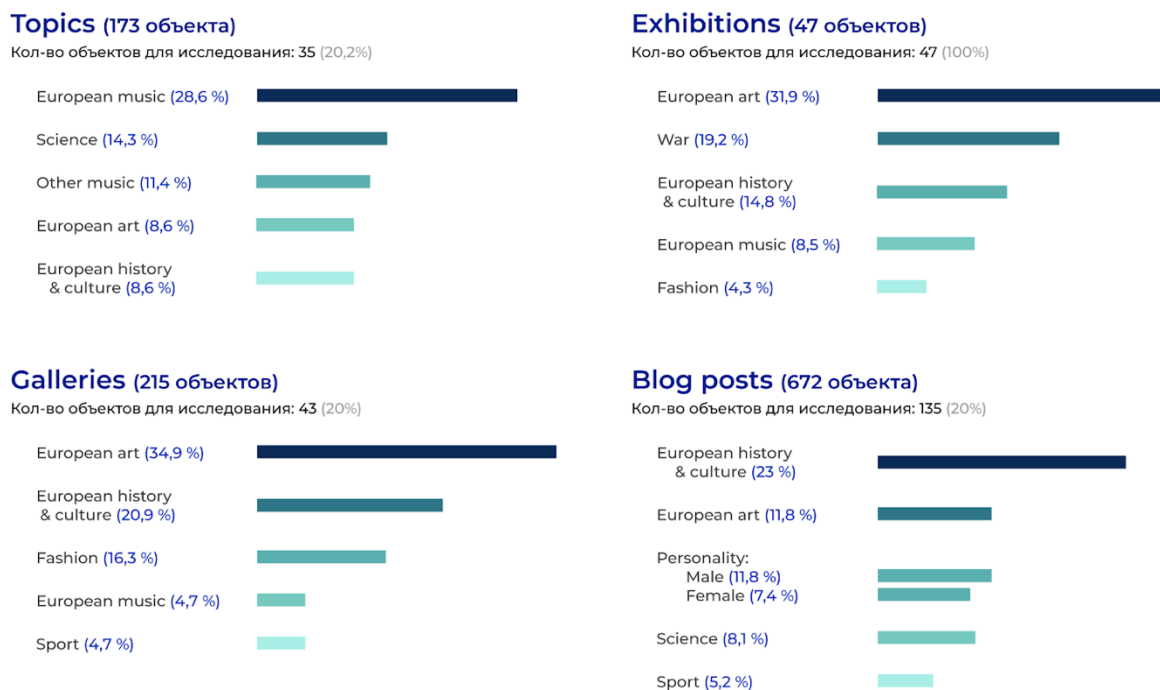


Рисунок 19 - Анализ содержания 4 главных разделов интерфейса Европеаны по 5 тематическим категориям. Автор: Арина Орлова

Инфографика, созданная на основании собранных данных, показывает, что в 4 крупнейших разделах большую часть объектов занимает европейское канонизированное искусство, музыка, записи по истории и культуре европейского общества.

С появлением новой версией сайта Европеаны, появилась возможность просмотра объектов естественно-научного и культурного наследия по вековым периодам, к которым принадлежат эти экспонаты.

В феврале 2022 года на основании собранных в таблицу данных была составлена диаграмма по каждому веку, демонстрирующая процентное соотношение 5 организаций, предоставивших максимальное количество материалов, к общему количеству учетных записей в каждом веке (см. рис. 20).

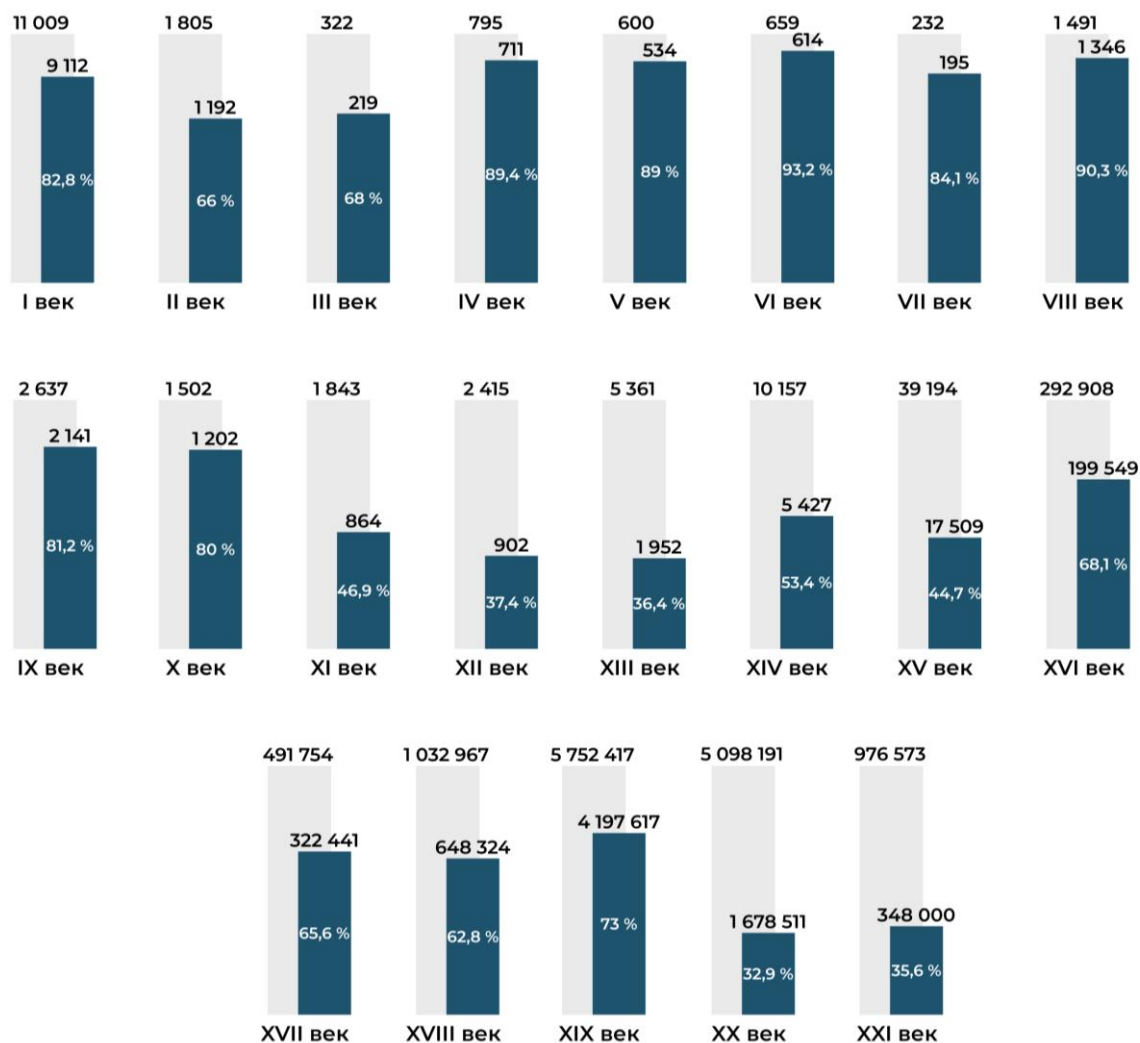


Рисунок 20 - Процентное распределение 5 учреждений культуры, представивших максимальное количество материалов, по отношению к общему количеству материалов в каждом веке. Автор: Арина Орлова

Также был проведен сравнительный анализ данных, показывающий количественное соотношение объектов культурного наследия за два периода: с I в. по XIX в. и период с XX в. по XXI в. (см. рис. 21).

Количество учетных записей по векам



Рисунок 21 - Сравнение количества учетных записей по двум временным периодам.

Автор: Арина Орлова

Оцифрованные материалы датируемые и тегируемые последними двумя веками составляют почти 1/2 от общего количества учетных записей в разделе временных периодов. Такого рода перекося может значительно влиять на формирование представлений пользователя о вековых периодах.

Была предпринята попытка проанализировать материалы размещаемые в каждом. Подводя итог этого анализа, следует отметить, что материалы размещаемые в этом разделе можно классифицировать на первичные и вторичные источники. Первые будут напрямую относиться к временному периоду и представлять существенную ценность для историко-культурного наследия, другие же, вторичные, лишь косвенно связаны с заявленным веком. Например, начиная с IX века большую часть материалов, если не всю, занимают сканированные страницы книг, причем датировка книг может не совпадать с вековым периодом, такие документы будут отнесены к первичным источникам.

На основании полученных в ходе анализа выборки данных можно констатировать, что контент цифровой библиотеки Европейцы не является тематически разнообразным и доминирующую часть записей в главных разделах составляет контент о европейском культурном наследии.

3.2 Анализ данных по сбалансированности материалов коллекций Европейцы

Подсчет количества учетных записей по каждой из 197 стран ООН с применением поисковой строки помог выявить лидеров по количеству учетных записей опубликованных каждой страной и составить диаграмму распределения коллекций стран по численности. По полученным данным был составлен рейтинг из 10 стран по наибольшему количеству учетных записей, предоставленных для агрегации (см. рис. 22).

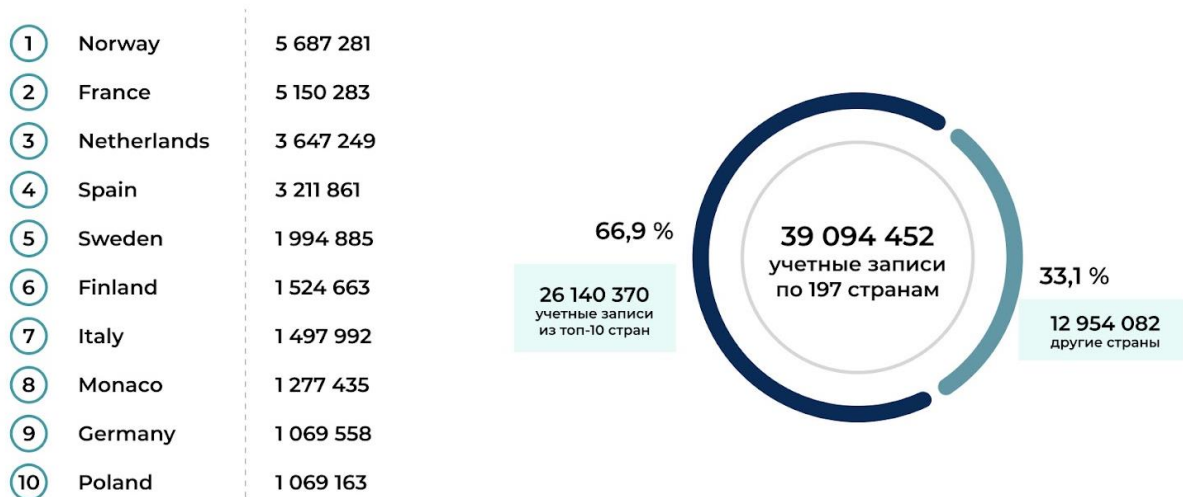


Рисунок 22 - Рейтинг 10 стран членов ООН по количеству учетных записей. Автор:
Арина Орлова

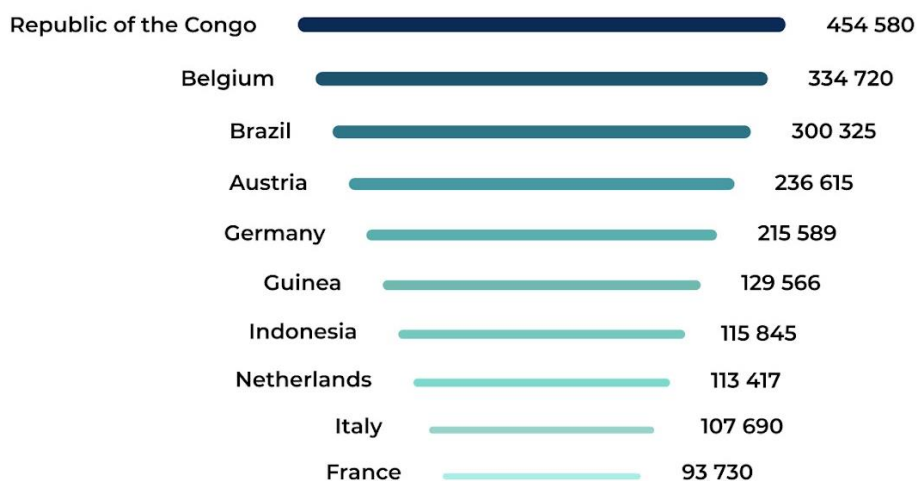


Рисунок 23 - Рейтинг из 10 стран в коллекции “Natural History” с указанием количества материалов, предоставленных для агрегации. Автор: Арина Орлова

Полученные результаты свидетельствуют о том, что 82,3% материалов были предоставлены для агрегации 5 крупными учреждениями культуры.

Данные о местоположении организации позволяют определить представленные коллекциями страны. На рисунке 24 показан рейтинг из 10 организаций и в каких странах они находятся, с указанием количества учетных записей.

Naturalis Biodiversity Center	Netherlands	4 533 393
The Trustees of the Natural History Museum, London	Great Britain	1 460 023
Meise Botanic Garden	Belgium	1 384 259
Royal Botanic Gardens, Kew	Great Britain	596 940
National Museum of Natural History	United States	501 471
The Royal Botanic Garden Edinburgh	Scotland	484 142
Berlin-Dahlem Botanical Garden and Botanical Museum	Germany	158 802
Natural History Museum, Vienna - Herbarium W	Austria	114 919
University of Trieste	Italy	101 316
Biologiezentrum der Oberoesterreichischen Landesmuseen	Austria	91 679

Рисунок 24 - Рейтинг из 10 организаций в коллекции “Natural History” с указанием страны размещения и количества материалов, предоставленных для агрегации в этой коллекции. Автор: Арина Орлова

Закономерным является тот факт, что главное место в этом рейтинге занимают Нидерланды (4 533 393), так как европейская цифровая библиотека базируется именно в этой стране.

На основании полученных данных 14 февраля 2022 года были составлены верхние диаграммы на рисунках 25 и 26, иллюстрирующие количество упоминаний учреждения культуры и стран в созданной ранее таблице. Но на момент 21 февраля 2022 года данные на сайте европейской цифровой библиотеки Европеаны об организациях и странах, предоставивших материалы отсортированные по временным периодам, изменились. Для понимания того, как изменились данные и как это может влиять на общую картину географического распределения материалов проводился повторный сбор данных и составлялись аналогичные диаграммы, которые представлены в нижней части рисунков ?? и ??. А также территориальное распределение организаций может помочь идентифицировать основных поставщиков цифрового контента.

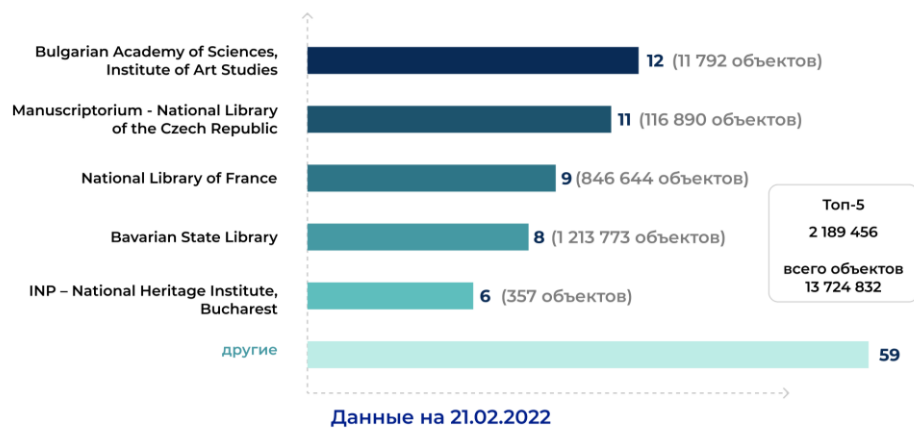
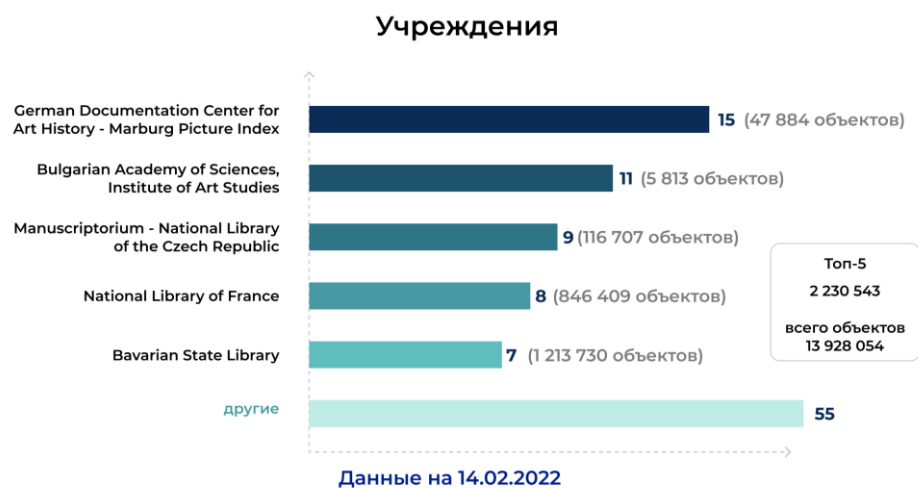
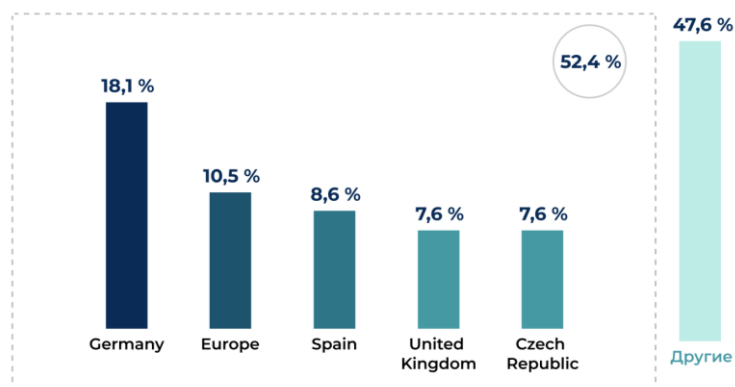
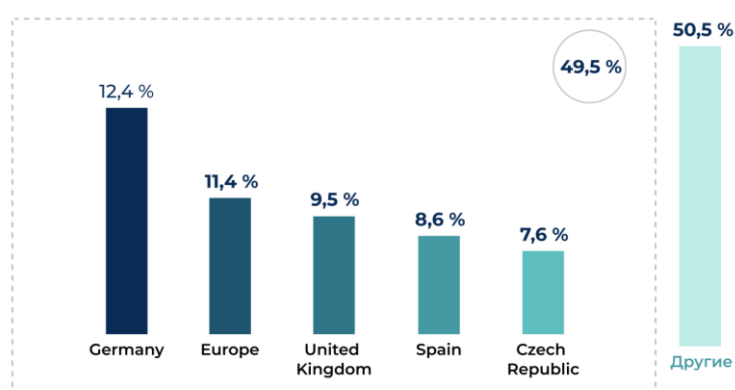


Рисунок 25 - Сравнение рейтингов из 5 учреждений культуры по количеству упоминаний в материалах Европеаны, отсортированных по векам. Автор: Арина Орлова

Страны



Данные на 14.02.2022



Данные на 21.02.2022

Рисунок 26 - Сравнение рейтингов из 5 стран по количеству упоминаний в материалах Европеана, отсортированных по векам. Автор: Арина Орлова

3.3 Обсуждение результатов исследования

Проведенное исследование позволяет сделать вывод о том, что разнообразие материалов историко-культурного наследия, которое демонстрирует европейская цифровая библиотека Европеана в наполнении главного интерфейса, на самом деле лишь поверхностное, и при более детальном изучении онлайн базы данных этого веб-ресурса проявляется недостаточная сбалансированность выборки внутри всех коллекций, что сказывается на репрезентативности тематических коллекций.

Еuropeана предоставляет доступ к материалам историко-культурного наследия 197 стран мира. Тем не менее можно констатировать

неравномерное распределение материалов коллекций разных стран. Из проведенного исследования видно, что наибольшая часть коллекций Европейцы представляет европейское искусство и включает материалы из крупных европейских музейных центров.

Несмотря на кажущееся разнообразие контента и его однородность по отношению к географии учреждений культуры, проведенный в ходе исследования анализ показывает, что существует большой снос данных по их географической классификации, а тематики контента не представляют собой однородное поле контента. Выявление данной проблемы важно потому, что снос данных по географическому признаку и прослеживаемая однородность контента, могут приводить к эпистемологическим пропускам [ВСТАВИТЬ ССЫЛКУ]. Вместе с тем проведенное исследование эксплицирует сегодняшнюю проблему агрегаторов коллекции культурного наследия и ставит задачу для последующего поиска закономерностей в выявленных пропусках.

ЗАКЛЮЧЕНИЕ

Цель настоящей работы - определить, является ли цифровая библиотека Европеана сбалансированным и разнообразным источником данных для гуманитарных исследований, используя количественный метод анализа.

В ходе исследовательской работы была продемонстрирована возможность применения понятий корпусной лингвистики, таких как репрезентативность, сбалансированность и разнообразие для оценки тематических коллекций цифровой библиотеки Европеана.

Были собраны данные о количестве основных коллекций Европеаны и количестве объектов в каждой. Были собраны и визуализированы данные о количестве материалов из 197 государств-членов ООН, размещаемых в Европеане. Собранные данные позволили выявить распределение материалов по коллекциям и определить главных стран-поставщиков европейского цифрового контента. Коллекция “Natural History” занимает первое место по числу учетных записей среди 13 основных коллекций, что составляет пятую часть содержимого Европеаны. При этом большую часть коллекции составляют биологические коллекции, сохранившиеся с колониального прошлого стран. Самая крупная коллекция естественно-научных материалов, принадлежащая Нидерландам, содержит 4 533 393 учетные записи. Также было выявлено, что Норвегия фигурирует большее количество раз по всем коллекциям, а именно, в результате поиска, обнаружилось 5 687 281 записей по этой стране.

Результаты исследования демонстрируют, что агрегатор оцифрованного культурно-исторического наследия Европеана не сбалансирован и не разнообразен, следовательно он не может выступать репрезентативным источником данных для.

Европейская цифровая библиотека представляется хранилищем огромного количества различного рода данных не обладающих сильным

тематическим разнообразием, пока этот веб-ресурс не может выступать в качестве первостепенного источника для объективных культурологических и гуманитарных исследований.

Будучи крупным проектом европейских стран, Европеана имеет все задатки ресурса, способного не только упростить, но и улучшить исследовательскую деятельность. Обеспечение тематических коллекций репрезентативными и сбалансированными массивами данных с разнообразным контентом позволит проводить анализ больших данных на основании объективных данных, предоставленных одним из крупнейших ныне существующих агрегаторов историко-культурного наследия.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Manovich, L. Cultural Data: Possibilities and Limitations of Working with Historical Cultural Data / L. Manovich // *Museum and Archive on the Move: Changing Cultural Institutions in the Digital Era*. – 2017. – С. 259–276.
2. Drucker, J. Is There a «Digital» Art History? / J. Drucker // *Visual Resources*. – 2013. – No 29. – С. 5–13.
3. Редькина, Н. С. Цифровые библиотеки: опыт США / Н. С. Редькина // *Библиосфера*. – 2008. – № 1. – С. 57–63.
4. Digital Humanities: гуманитарные науки в цифровую эпоху: монография / У. С. Захарова, Г. В. Можаяева, П. Н. Можаяева, Ж. А. Рожнева, В. А. Сербин, А. А. Хаминаова; под редакцией Г. В. Можаяевой. – Томск: Изд-во Том. ун-та, 2016. – 120 с. – ISBN 978-5-7511-2468-7.
5. Lynch, C. A. Digital collections, digital libraries and the digitization of cultural heritage information / C. A. Lynch // *First Monday*. – 2002. – Vol. 7, Iss. 5.
6. Горный, Е. Развитие электронных библиотек: мировой и российский опыт, проблемы, перспективы / Е. Горный, К. Вигурский // *Интернет и российское общество* / редактор И. Семенова; Моск. Центр Карнеги. – М.: Гендальф, 2002. – 279 с.
7. Bharathkumar, V. Digital Transformation: Need for the Transformation of Libraries into Digital Libraries / V. Bharathkumar // *International Conference on Digital Technologies and Transformation in Academic Libraries* / J.R.D. Tata Memorial Library, Indian Institute of Science. – Bengaluru, 2022.
8. Воронцова, Ю. В. Проблемы управления цифровыми библиотеками / Ю. В. Воронцова, А. Д. Агибалова // *E-Management*. – 2021. – Т. 4, № 2. – С. 51–57.
9. Марчук, А. Г. Особенности построения цифровых библиотек со связанным контентом / А. Г. Марчук, П. А. Марчук // *Электронные*

библиотеки: перспективные методы и технологии, электронные коллекции: Сб.трудов / XII Всеросс. научн. Конф. RCDL'2010, Казань, Россия 13-17 октября 2010 г. - Казань: Казан. ун-т, 2010 . – 2020. – № 9. – С. 19–23.

10. Боярский, К. К. Введение в компьютерную лингвистику / К. К. Боярский. – СПб: НИУ ИТМО, 2014. – 72 с.

11. Рыков, В. В. Корпус текстов как реализация объектно-ориентированной парадигмы / В. В. Рыков // Труды Международного семинара «Диалог-2002». – М. : Наука, 2002. – С. 124–129.

12. Плунгян, В. А. Почему современная лингвистика должна быть лингвистикой корпусов [Электронный ресурс] / В. А. Плунгян // Лекция, прочитанная в рамках проекта «Публичные лекции». – 2009. – Режим доступа: <http://polit.ru/article/2009/10/23/corpus/>

13. Елизаров, А. М. Цифровые платформы и цифровые научные библиотеки / А. М. Елизаров, Е. К. Липачёв // International Journal of Open Information Technologies. – 2020. – Vol. 8, № 11. – P. 80–88.

14. Захаров, В. П. Корпусная лингвистика: учебник / В. П. Захаров, С. Ю. Богданова. – 3-е изд., перераб. – СПб.: Изд-во С.-Петербур. ун-та, 2020. – 234 с. – ISBN 978-5-288-05997-1.

15. Прихожев, П. В. Цифровая библиотека «Европеана» как проект доступа к объектам европейского культурного наследия / П. В. Прихожев // Культура: теория и практика. – 2020. – № 3 (36). – С. 80–82.

16. Manovich, L. Cultural Analytics / L. Manovich. – Cambridge, MA: The MIT Press, 2020. – P. 318. – ISBN 978-0-2620-3710-5.

17. Редькина, Н. С. Европеана: цифровое культурное наследие Европы / Н. С. Редькина // Ученые записки (Алтайская государственная академия культуры и искусств).. – 2020. – № 2 (24). – С. 89–96.

18. Valtysson, B. EUROPEANA: The digital construction of Europe's collective memory / B. Valtysson // Information, Communication & Society. – 2012. – Vol. 15, Iss. 2. – P. 151–170.

19. Bergen, N. "Everything Is Perfect, and We Have No Problems": Detecting and Limiting Social Desirability Bias in Qualitative Research / N. Bergen, R. Labonte // Sage Journals. – 2019. – Vol. 30, Iss. 5. – P. 783–792.

20. Кижнер, И. А. Цифровые технологии культурного наследования в современном обществе : специальность 24.00.01 "Теория и история культуры" : диссертация на соискание ученой степени кандидата культурологии / Кижнер Инна Александровна ; Сибирский федеральный университет. – Красноярск, 2021. – 182 с.


Федеральное государственное автономное
образовательное учреждение высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Гуманитарный институт

Кафедра информационных технологий в креативных и культурных индустриях

УТВЕРЖДАЮ

И. О. заведующего кафедрой

 А. В. Усачёв

подпись

« 5 » 07 2022 г.

БАКАЛАВРСКАЯ РАБОТА

09.03.03.14 Прикладная информатика в области искусств и гуманитарных наук

Оценка репрезентативности тематических коллекций цифровой библиотеки
Европеана

Руководитель



подпись, дата

05.07.2022

д-р филос. наук

Р.П. Мусат

Выпускник



подпись, дата

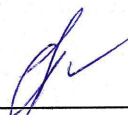
05.07.22

А.П. Орлова

Красноярск 2022

Продолжение титульного листа бакалаврской работы по теме Оценка репрезентативности тематических коллекций цифровой библиотеки Европеана

Нормоконтролер



подпись, дата

Е.Р. Брюханова