

Историческая информатика

Правильная ссылка на статью:

Антопольский А.Б., Володин А.Ю. — Справочно-информационная система по цифровой гуманитаристике: опыт описания интернет-ресурсов российских архивов // Историческая информатика. – 2022. – № 2. – С. 50 - 66. DOI: 10.7256/2585-7797.2022.2.38236 EDN: HOVPGY URL: https://nbpublish.com/library_read_article.php?id=38236

Справочно-информационная система по цифровой гуманитаристике: опыт описания интернет-ресурсов российских архивов

Антопольский Александр Борисович

кандидат филологических наук, доктор технических наук

Главный научный сотрудник, ИНИОН РАН

117218, Россия, г. Москва, ул. Кржижановского, 15

✉ ale5695@yandex.ru



Володин Андрей Юрьевич

кандидат исторических наук

доцент кафедры исторической информатики МГУ имени М.В. Ломоносова, руководитель Digital Humanities Research Institute Сибирского федерального университета

119991, Россия, г. Москва, Ломоносовский проспект, 27к4, оф. Г-423

✉ volodin@hist.msu.ru



[Статья из рубрики "Цифровые ресурсы"](#)

DOI:

10.7256/2585-7797.2022.2.38236

EDN:

HOVPGY

Дата направления статьи в редакцию:

08-06-2022

Аннотация: Статья содержит описание результатов инвентаризации Интернет-ресурсов российских архивов, проведенной в целях создания справочно-информационной системы по цифровой гуманитаристике (СИСЦГ). Определяются объекты, образующие в совокупности инфосферу этой дисциплины. Рассматривается проблема детализации уровня ресурсов при их учете и каталогизации. Обсуждаются критерии отбора специфических ресурсов, интересных для пользователя. Приводится структура базы данных проектируемой системы. Обсуждаются вопросы применения языков для индексирования информационных объектов. Приводятся количественные характеристики

проведенной инвентаризации. СИС ЦГ должна обеспечивать повышение видимости, цитируемости и повторного использования создаваемых в России объектов цифровой гуманитаристики. В данном исследовании были изучены представленные в Интернете ресурсы федеральных архивов, государственных архивов субъектов РФ, а также в небольшом количестве - ведомственные архивы, если они имеют социогуманитарное содержание. Органы управления архивами субъектов РФ включались в БД, если у них есть собственный сайт, на котором размещены ресурсы, представляющие интерес. Архивы, имеющие доминирующее техническое или естественно-научное содержание, в исследование не включались. Также на данном этапе не рассматривались ресурсы муниципальных архивов. В ходе исследования были определены также основные методические проблемы учета и каталогизации интернет-ресурсов, которые необходимо решить.

Ключевые слова: Российские архивы, интернет-ресурсы, инвентаризация, каталогизация, индексирование, цифровая гуманитаристика, интернет-эвристика, интернетика, информационный поиск, информационная система

При поддержке программы стратегического академического лидерства «Приоритет-2030».

Введение

ИНИОН РАН и Институт цифровых гуманитарных исследований (DHRI) Сибирского федерального университета заключили соглашение о совместном создании и поддержке Справочно-информационной системы по цифровой гуманитаристике (СИС ЦГ). Эта система должна содержать информацию о созданных в мире и России объектах ЦГ — ресурсах, сервисах, нормативах, инструментах, проектах — в целях их повторного использования и исключения дублирования при разработке. Предполагается, что подобная СИС ЦГ может быть использована:

- при разработке программ НИР;
- при подготовке и экспертизе заявок на гранты, проекты, темы НИР;
- для оценки результатов НИР, а также учреждений и научных подразделений;
- для координации работ по оцифровке объектах ЦГ;
- для информационного обеспечения научной и образовательной деятельности по применению компьютерных технологий в области гуманитарных и общественных наук.

С И С ЦГ должна обеспечивать повышение видимости, цитируемости и повторного использования создаваемых в России объектов ЦГ.

Перспективной целью СИС ЦГ должно стать формирование основы для решения задачи определения оптимального уровня интеграции ресурсов и коллаборации деятельности в области цифровой гуманитаристики и шире, при создании инфраструктуры цифровой науки в России.

Тематические границы инфосферы, которая должна быть отражена в СИС ЦГ, еще обсуждаются и будут уточнены в процессе проектирования и практического наполнения контента СИС ЦГ. Частично эти вопросы рассмотрены в работе [\[1\]](#). Однако очевидно, что в институциональном разрезе в инфосферу ЦГ входят ресурсы библиотек, архивов и

музеев. Поэтому именно эти ресурсы, причем в основном социогуманитарной направленности, должны быть исследованы в первую очередь.

Результаты представленного исследования должны быть использованы при формировании контента СИС ЦГ.

Цель исследования

Цель настоящего исследования, если ее сформулировать коротко – проведение инвентаризации ресурсов российских архивов, представленных в Интернете, и формирование базы данных описаний этих ресурсов. Эта база данных может рассматриваться как пилотная версия контента проектируемой системы. Очевидно, при этом нужно провести систематизацию ресурсов и определение структуры и состава базы данных.

История вопроса

Исследование интернет-ресурсов отечественных архивов имеет свою историю. Особый вклад в него внесла Е.В.Боброва, которая провела несколько исследований этих ресурсов, в том числе в 2005 г. [2], а также в 2007 г. и в 2018 г. [3,4]. Е.В.Боброва подробно проанализировала состояние архивных сайтов, как их контента, так и технологических особенностей, а также, что особенно важно, пользовательских аспектов. В ее работах приведена также статистика состояния интернет-ресурсов и динамики их развития.

Многие важные вопросы организации архивных интернет-ресурсов рассмотрены в фундаментальных монографиях Ю.Ю. Юмашевой [5,6]. Первая из них, посвященная информатизации архивного дела в России в 1991-2015 гг., рассматривает в широком контексте вопросы создания тематических архивных баз данных и информационных ресурсов, историю архивного сайтостроения и др. Вторая монография Ю.Ю. Юмашевой посвящена проблемам цифровой трансформации традиционных аудиовизуальных архивов, переводу аудио-, фото-, кинодокументов в цифровой вид, созданию электронных ресурсов на основе созданных копий и их представлению в сети Интернет. В основе монографии лежит созданная автором база данных, в которой представлены доступные в мировой паутине информационные ресурсы, основанные на переведенных в цифровой вид библиографических каталогах, научно-справочном аппарате архивов и учетных документах музеев, а также электронных копиях объектов.

Отметим также работу И.Н. Киселева [7], посвященную анализу размещенных в Сети на сайтах государственных и муниципальных архивов, а также уполномоченных органов исполнительной власти субъектов Российской Федерации в области архивного дела и на портале «Архивы России» справочников по архивным документам, степени их доступности для пользователей Интернета, форматов и некоторых иных свойств. Все данные собраны автором путем сплошного просмотра сайтов по состоянию на октябрь–ноябрь 2017 г.

Интересный опыт систематизации историко-ориентированных информационных систем был предпринят командой пермских историков во главе с С.И. Корниенко [8]. Обобщенные результаты представлены в коллективной монографии «Исторические информационные системы: теория и практика», в которой историко-ориентированные информационные системы рассматриваются как значимый элемент информационной среды гуманитарных наук [9].

Особенности международного опыта описания ресурсов и инструментов, применяемых в цифровой гуманитаристике, фактурно представлены в коллективной статье, посвященной слиянию каталога DiRT (Digital Research Tools) с порталом TAPoR (Text Analysis Portal for Research) [\[10\]](#).

Однако цитированные авторы не касались методических вопросов учета и каталогизации интернет-ресурсов, что стало содержанием настоящего исследования.

Сфера исследования

В данном исследовании были изучены представленные в Интернете ресурсы федеральных архивов, государственных архивов субъектов РФ, а также в небольшом количестве - ведомственные архивы, если они имеют социогуманитарное содержание. Органы управления архивами субъектов РФ включались в БД, если у них есть собственный сайт, на котором размещены ресурсы, представляющие интерес. Таковых оказалось примерно половина из числа субъектов РФ.

Архивы, имеющие доминирующее техническое или естественно-научное содержание, в исследование не включались. Также на данном этапе не рассматривались ресурсы муниципальных архивов.

Основным источником информации о составе государственных архивов, а также органов управления архивным делом, послужил портал Архивы России [\[11\]](#), где имеется специальный раздел «Органы управления архивным делом и архивы субъектов Российской Федерации».

Объекты исследования

К объектам, предназначенным для включения в СИС ЦГ, в принципе относятся следующие типы информационных объектов:

- *Институции* – т.е. органы управления, организации, или подразделения учреждений, имеющие собственные интернет-ресурсы.
- *Ресурсы* – т.е. массивы документов или данных (сайты, разделы или страницы сайтов, базы данных, наборы данных, коллекции, документов или их описаний, каталоги, справочники и др.).
- *Сервисы* – доступные автоматизированные процессы для реализации задач пользователя.
- *Нормативы* – стандарты, методики, языки индексирования и другие средства создания ресурсов или реализации сервисов.
- *Инструменты* – программные продукты, используемые для создания ресурсов или реализации сервисов.
- *Проекты* – объекты, которые нельзя отнести к перечисленным видам, но представляющие интерес с точки зрения целей СИС ЦГ.

Эти типы были выделены на основе анализа зарубежного опыта, описанного в работе [\[1\]](#). При исследовании архивных ресурсов практически анализировались только три типа объектов: институции, ресурсы и сервисы. Нормативы и инструменты рассматривались на других этапах настоящего исследования. Что же касается проектов, то применительно к архивной деятельности различение проектов и ресурсов оказалось затруднительным и

нерелевантным.

Перечень видов институций, включенных в рассмотрение, был перечислен выше.

Что касается ресурсов и сервисов, то в ходе исследования решались две методические проблемы. Во-первых, это задача определения оптимального уровня детализации для выделения ресурса как объекта учета. Во-вторых, это различение типовых ресурсов и сервисов, включение которых в СИС ЦГ было признано нецелесообразным, от специфических, представляющих интерес для внешнего пользователя с точки зрения целей СИС ЦГ. Далее эти проблемы рассматриваются подробнее.

Проблема уровня детализации ресурсов

Определение оптимального уровня детализации объектов учета является одной из наиболее сложных методических проблем при создании систем навигации по информационному пространству. Например, в руководстве по применению метаданных для известного репозитория лингвистических данных OLAC [\[12\]](#) эта проблема занимает одно из центральных мест.

Напомним, что ресурсами мы называем поименованные (выделенные) массивы документов/данных, которые и являются объектами учета и поиска. Поскольку речь идет о сетевых ресурсах, идентификатором массива обычно выступает сетевой адрес или URL. Другие способы идентификации массива, такие как DOI, пока распространены недостаточно и в качестве инструмента отбора быть использованы не могут.

С одной стороны, детализация выделяемых массивов не должна быть слишком подробной, чтобы схема навигации оставалась обозримой и массивы однородных данных имели общее наименование. Поэтому навигация обычно не опускается до уровня отдельного документа или датасета.

С другой стороны, необходимо дать представление о составе сложных неоднородных информационных объектов, таких как сайты или порталы.

С третьей стороны, при выборе уровня детализации при формировании ресурса разработчики могли руководствоваться самыми разными представлениями или даже техническими причинами.

Рассмотрим пример. На сайте Национального архива республики Саха (Якутия) [\[13\]](#) имеется раздел под названием *Информационно-познавательная рубрика по истории Якутии*, который представляет интерес для СИС ЦГ. Этот раздел имеет следующую структуру.

- Древняя и средневековая Якутия
- Якуты
- Родовой состав якутов
- Якутия в составе Российского государства
 - Вхождение Якутии в состав России
 - Якутск и русские географические открытия XVII века
 - Ясак
 - г. Якутск
 - г. Олекминск
 - г. Вилуйск
 - г. Верхоянск

- г. Среднеколымск
- Якутия в XVIII-XIX вв.
- Якутия в начале XX века. (1900-1917 гг.)
- Февральская и Октябрьская революции
- Гражданская война
- Якутия в годы Великой Отечественной войны 1941-1945 гг.
- 1922-1929 гг.
- 1930-1935 гг.
- 1936-1941 гг.
- Якутия в постсоветский период. 1992-1999 гг.

Однако этот раздел целиком не имеет своего идентификатора в виде адреса, что обусловлено техническими причинами. Также нет отдельного адреса у рубрик, имеющих второй уровень, таких как *Якутия в составе Российского государства*. Поэтому возникает дилемма: или ограничиться уровнем всего сайта, или включать в БД СИС ЦГ как отдельные ресурсы рубрики на всех уровнях.

Другой пример. На большинстве сайтов архивов представлен научно-справочный аппарат данного архива. В качестве примера приведем раздел *Справочно-поисковые средства* сайта Национального архива Республики Татарстан [\[14\]](#).

ПУТЕВОДИТЕЛИ И СПРАВОЧНИКИ ПО ФОНДАМ

Центральный государственный архив историко-политической документации Республики Татарстан. Путеводитель.

Справочник документов по личному составу по фондам ГБУ «Государственный архив Республики Татарстан» (2011-2018 гг.)

Государственный архив документов по личному составу Республики Татарстан: Краткий справочник (2005-2010 гг.)

Краткий справочник по фондам Национального архива Республики Татарстан (2010-2014 гг.).

Краткий справочник по фондам Национального архива Республики Татарстан (2005-2009 гг.)

СПРАВКИ-ОРИЕНТИРОВКИ

Формулярные списки

Похозяйственные книги

Справка-ориентировка о переписных листах по проведению первой всеобщей переписи населения Российской империи 1897 года

Список эвакогоспиталей, находившихся на территории Татарской АССР в период Великой Отечественной войны 1941-1945 гг.

УКАЗАТЕЛИ

Указатель чертежей и планов г. Казани (том 1)

Указатель чертежей и планов г. Казани (том 2)

Указатель чертежей и планов г. Казани (том 3)

Межфондовый указатель клировых ведомостей

Межфондовый указатель карт и планов населенных пунктов, земельных и лесных участков...

В этом случае напротив, как раздел целиком, так и массивы/документы второго и третьего уровней имеют самостоятельные наименования и адреса-идентификаторы и могут быть включены в БД СИС ЦГ как самостоятельные ресурсы. Однако возникает вопрос о целесообразности подобной детализации для задач СИС ЦГ.

Очевидно, что вопрос об оптимальном уровне детализации при описании ресурсов будет окончательно решен в ходе эксплуатации СИС ЦГ с учетом интереса пользователей. Пока для описанного исследования применялся принцип наибольшего уровня идентификации среди однородных данных, имеющих общий идентификатор. Т.е. для данного примера раздел *Справочно-поисковые средства* был включен в БД целиком. Предполагается, что при практическом функционировании СИС ЦГ оптимальный уровень детализации может быть изменен в ту или другую сторону.

Проблема типовых и специфических ресурсов и сервисов

Легко видеть, что сайты российских архивов в общем случае весьма насыщены разнообразной информацией. Однако в какой степени эта информация будет полезна для пользователей СИС ЦГ, на данном этапе неясно.

Основным критерием, для включения ресурсов в БД СИС ЦГ был принят следующий. Отбирались ресурсы, специфические для данного архива, а ресурсы, типические или стандартные для многих архивов, не включались. При этом, как было указано выше, отдельные документы не включались, если они не представляли собой перечни данных, такие как путеводители, указатели, библиографии и проч.

Важное место среди отобранных ресурсов составили перечни проектов, выставок, услуг, ресурсов и других информационных объектов. Эти перечни должны служить основой для возможного расширения контента СИС ЦГ.

Конечно, включались только цифровые ресурсы, но представленные в любых форматах, как текстовых, так и в графических. Описания традиционных ресурсов, в частности библиотек архивов, не включались, если они не имели электронных услуг или ресурсов

Особый вопрос о включении в БД СИС сервисов (услуг). Напомним, что задача СИС ЦГ - отражать прежде всего цифровую инфосферу цифровой гуманитаристики. С другой стороны, большинство архивных услуг сервисов является типовыми для архивной отрасли. Поэтому в БД были включены в основном услуги, выполняемые с помощью использования БД, а также перечни услуг.

Следует также упомянуть о конференциях, семинарах и других мероприятиях, проводимых архивами. На данном этапе эта информация не включалась, потому что она зачастую носит непостоянный характер. Кроме того, существуют специальные ресурсы, на которых отражена информация о мероприятиях, поэтому прямое их дублирование выглядит нецелесообразным. Однако вопрос об отражении в СИС ЦГ научных мероприятий требует отдельного обсуждения.

В результате получился следующий примерный состав ресурсов, отобранных с сайтов архивов для включения в СИС ЦГ. Перечень представлен ключевыми словами, отражающими вид информационных ресурсов.

- Архивные коллекции
- Аудиовизуальные ресурсы
- Базы данных
- Библиографии
- Виртуальные ресурсы
- Выставки
- Генеалогия
- Географические информационные системы
- Геоданные
- Интернет-проекты
- Информационно-поисковые системы
- Каталоги
- Музейные коллекции
- Мультимедийные ресурсы
- Наборы открытых данных
- Научно-справочный аппарат
- Описи
- Перечни (списки) лиц,
- Перечни проектов
- Перечни ресурсов
- Периодические издания
- Персональные коллекции
- Порталы широкого профиля
- Путеводители
- Словари
- Справочники
- Тематические коллекции
- Тематические подборки
- Указатели
- Фотогалереи
- Хронологические указатели и календари
- Цифровые издания и публикации архивов
- Электронные библиотеки
- Энциклопедии

Для полноты картины приведем также примерный перечень типовых разделов архивных сайтов, которые как правило, не включались в базу данных СИС ЦГ.

- Структура архива
- История архива
- Состав документов архива
- Новые поступления
- Источники комплектования архива
- Научный совет архива
- Экспертно-проверочная комиссия архива
- Партнёры архива
- Доска почёта архива
- Отчеты архива
- Нормативные документы архива
- Противодействие коррупции
- Мероприятия

- Читальный зал

Структура базы данных СИС ЦГ

Исходным вариантом базы данных является структура, состоящая из следующих полей представления объекта:

- о Акроним объекта, который выполняет также функции связи ресурса с владельцем;
- о Полное наименование объекта на языке оригинала;
- о Перевод наименования на русский язык или наименование объекта на русском языке;
- о Веб-адрес объекта;
- о Тип объекта (в соответствии с разделом *Объекты исследования* .);
- о Вид или функциональность объекта (в настоящее время 1-3 ключевых слова, разделенных запятой);
- о Рубрика модернизированного ГРНТИ (в настоящее время 1-3 кода рубрик, разделенных запятой).

Эта структура уже опробована при создании некоторых навигаторов, функционально аналогичных СИС ЦГ [\[15,16\]](#), и оказалась вполне работоспособной.

Здесь кратко обсудим некоторые проблемы, вытекающие из данной структуры. Прежде всего, это проблема идентификатора. Идея состояла в том, что акроним институции должен служить ключом для связи ресурсов, сервисов и проектов, принадлежащих этой институции с владельцем. Это позволит показывать при поиске все ресурсы данной институции, что бывает очень удобно.

В принципе в большинстве случаев это вполне реализуемо и принадлежность ресурса владельцу определяются однозначно. Однако встречаются случаи, когда возникают некоторые сложности.

Во-первых, не всегда возможно использовать в качестве акронимов официальные аббревиатуры организации, поскольку они могут совпадать, потому приходится использовать искусственные акронимы.

Во-вторых, некоторые ресурсы могут принадлежать нескольким владельцам, т.е. отношение владения может не быть однозначным.

В-третьих, существуют ресурсы, для которых институцию-владельца не удастся установить.

Очевидно, что при практической реализации СИС ЦГ следует предусмотреть технологические способы решения этих проблем.

Языки индексирования

Далее рассмотрим возможности и перспективы использования для СИС ЦГ некоторых средств индексирования, исходя из основной задачи этой системы, которую технологически можно определить как поиск аналогов для проектов, ресурсов и сервисов в сфере цифровой гуманитаристики. Более подробно средства индексирования для цифровой гуманитаристики в целом проанализированы в работе [\[17\]](#), а здесь

рассмотрим применение языков индексирования для описания архивных ресурсов.

В описанном исследовании, кроме общего разделения информационных объектов на типы использовались 1) индексирование объектов для определения вида объекта при помощи ключевых слов; 2) определение тематики при помощи модернизированного рубрикатора ГРНТИ.

Ключевые слова для определения вида объекта *Информационные ресурсы* в основном приведены выше, в разделе *Проблема типовых и специфических ресурсов и сервисов*

Для определения вида объекта *Институция* были использованы следующие ключевые слова

- Архив субъекта РФ
- Ассоциация
- Библиотеки федеральных органов исполнительной власти
- Библиотеки отраслевых академий
- Депозитарный архив
- НИИ
- Орган управления архивами
- Отраслевые академии
- Федеральные архивы

Для определения вида объекта *Сервис* использовались ключевые слова:

- Информационное обслуживание
- ИПС
- Генеалогия
- Перечень услуг

Столь скромный перечень видов сервисов, был объяснен выше: сервисы были включены в БД в очень ограниченном объеме.

В целом определение видов информационных объектов при помощи открытого перечня ключевых слов, которые также могут быть использованы в качестве поисковых тегов, не вызывает проблем.

Сложнее выглядит ситуация с определением тематики ресурсов с использованием ГРНТИ. В цит. работе [\[1\]](#) описывалось применение модернизированного ГРНТИ для индексирования зарубежных проектов по цифровой гуманитаристике. Этот опыт может быть признан достаточно успешным, вероятно, от того что зарубежные проекты имели очевидную привязку к определенной тематике и/или к определенным информационным технологиям. В отношении архивных ресурсов дело обстоит не так. За исключением архивных коллекций, относящихся к военной истории, и небольшого количества тематических коллекций по другим отраслям или эпохам, большинство архивных ресурсов тематически не окрашено и для них приходилось использовать общую рубрику

13.71 Архивное дело. Архивоведение. Напомним деление этой рубрики в ГРНТИ, причем рубрика 13.71.85 Цифровые архивы была введена нами в порядке модернизации ГРНТИ.

13.71 Архивное дело. Архивоведение

13.71.01 Общие вопросы

13.71.07 Теория и практика архивного дела

13.71.09 История архивного дела

13.71.15 Организация и управление в области архивного дела

13.71.85 Цифровые архивы

13.71.91 Архивное дело в отдельных странах

13.71.96 Естественнаучное и техническое обеспечение архивного дела

Очевидно, что такое деление не позволяет тематически структурировать большинство архивных ресурсов, состав которых показан выше. При этом следует опираться на общий методический принцип использования классификации для навигации. Согласно этому принципу рубрикация должна быть такой, чтобы количество объектов в каждой рубрике было обозримым для интеллектуального просмотра, т.е. не превышать 30-40 пунктов.

Из этого следует, что для задач СИС ЦГприменительно к архивным ресурсам нужно или разработать принципиально другую схему, или отказаться от тематической классификации, ограничившись предметными тегами.

Основные количественные результаты исследования

Всего в ходе исследования сайтов российской архивной отрасли с учетом описанных выше ограничений и методических принципов было выделено св. 1.5 тыс. информационных объектов, в том числе ресурсов - 1250, институций - 256, сервисов - 28. На табл.1 приводятся количественные данные по видам объектов.

Табл.1 Количество информационных объектов

Тип объекта	Вид объекта	Количество объектов
Информационные ресурсы	Архивные коллекции	169
Информационные ресурсы	Аудио	5
Информационные ресурсы	Библиография	62
Информационные ресурсы	Видео	24
Информационные ресурсы	Виртуальная реальность	6
Информационные ресурсы	Выставки	113
Информационные ресурсы	ГИС	9
Информационные ресурсы	Данные. Наборы	20

Информационные ресурсы	Данные, ресурсы	55
Информационные ресурсы	Изображения	10
Информационные ресурсы	ИПС	108
Информационные ресурсы	Медиа	10
Информационные ресурсы	Музейные коллекции	12
Информационные ресурсы	Очерк	5
Информационные ресурсы	Памятники	3
Информационные ресурсы	Перечень лиц	29
Информационные ресурсы	Перечень проектов	12
Информационные ресурсы	Перечень ресурсов	19
Информационные ресурсы	Периодика	11
Информационные ресурсы	Персональные коллекции	1
Информационные ресурсы	Портал	7
Информационные ресурсы	Словари	1
Информационные ресурсы	Справочники	364
Информационные ресурсы	Фото	33
Информационные ресурсы	Хронологии	33
Информационные ресурсы	Цифровые издания	110
Информационные ресурсы	Электронные библиотеки	14
Информационные ресурсы	Энциклопедии	2
Институты	Архив субъекта РФ	166
Институты	Архив ФОИВ	5
Институты	Ассоциация	4
Институты	Библиотеки	4
Институты	Депозитарные архивы	4
Институты	НИИ	5
Институты	Орган управления архивами	48
Институты	Федеральные архивы	20
Сервисы	Генеалогия	3
Сервисы	Информационное обслуживание	6

Сервисы	ИПС	4
Сервисы	Перечень услуг	15

Заключение

Проведенное исследование позволило достаточно наглядно представить структуру, количественные и качественные характеристики интернет-ресурсов, размещенных на сайтах российских архивов. Конечно, это не окончательные данные: в процессе реального наполнения и эксплуатации СИС ЦГ и состав информационных объектов и средства их описания могут и должны изменяться и развиваться. В ходе исследования были определены также основные методические проблемы учета и каталогизации интернет-ресурсов, которые необходимо решить. Выше мы указывали на необходимость решения проблем детализации ресурсов или тематической классификации. Очевидно, важную роль в способе организации данных сыграет и интерфейс СИС ЦГ, который еще предстоит разработать. В ближайшем будущем информационная система будет доступна онлайн для пилотного использования желающими. Нужно также иметь в виду, что ресурсы архивов — это только часть контента проектируемой системы по цифровой гуманитаристике.

Библиография

1. Антопольский А.Б. Инфосфера цифровой гуманитаристики: опыт анализа // Информационные ресурсы России. 2022. № 1. С. 30-38. DOI: 10.52815/0204-3653_2022_01185_30
2. Боброва Е. В. Архивы via Интернет // Новое литературное обозрение. – 2005. – № 4(74). – С.507-520.
3. Боброва Е. В. Российские архивные сайты: опыт, проблемы, перспективы развития // Отечественные архивы. – 2007. – № 5. – С. 128-130.
4. Боброва Е. В. Сайты российских архивов: взгляд пользователя // Отечественные архивы. – 2018. – № 6. – С. 28-38.
5. Юмашева Ю.Ю. Информатизация архивного дела в Российской Федерации (1991-2015 гг.). Научные исследования в области применения информационных технологий. М.-Берлин: Директ-Медиа, 2016. 360 с.
6. Юмашева Ю.Ю. Цифровая трансформация аудиовизуальных архивов. Аудиовизуальные архивы онлайн. М.-Берлин: Директ-Медиа, 2016. 599 с.
7. Киселев И.Н. НСА к архивным документам в Рунете // Отечественные архивы. 2018. № 1. С. 33–39.
8. Историко-ориентированные информационные системы (2013-2020, Корниенко С.И., Гагарина Д.А., Поврозник Н.Г. и др.). URL: <http://digitalhistory.ru/>
9. Корниенко С.И., Гагарина Д.А., Поврозник Н.Г. Исторические информационные системы: теория и практика. М.: ИД НИУ ВШЭ, 2021. 232 с.
10. Grant, K., Dombrowski, Q., Ranaweera, K., Rodriguez-Arenas, O., Sinclair, S., & Rockwell, G. (2020). Absorbing DiRT: Tool Directories in the Digital Age. Digital Studies/le Champ Numérique, 10(1), None. DOI: <http://doi.org/10.16995/dscn.325>
11. Архивы России. Органы управления архивным делом и архивы субъектов Российской Федерации. URL: <https://rusarchives.ru/state/list> (дата обращения 30.05.2022)
12. OLAC Metadata Usage Guidelines. – URL: <http://www.language-archives.org/NOTE/usage.html> (дата обращения: 30.05.2022).
13. Национальный архив Республики Саха (Якутия)– URL: <http://archivesakha.ru/> (дата

обращения: 30.05.2022)

14. Национальный архив Республики Татарстан – URL:

https://ga.tatarstan.ru/rus/file/pub/pub_2751655.docx (дата обращения: 30.05.2022)

15. Навигатор информационных ресурсов по общественным наукам – URL:

<http://niron.inion.ru/> (дата обращения: 30.05.2022)

16. Навигатор информационных ресурсов по языкознанию – URL: <http://niryaz.inion.ru/> (дата обращения: 30.05.2022)

17. Антопольский А.Б. Языки индексирования для цифровой гуманитаристики Научно-техническая информация ISSN 0548-0027. Сер. 2 Информационные процессы и системы. 2022 № 1 с.1-9 DOI: 10.36535/0548-0027-2022-01-1

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Статья посвящена проблеме разработки информационной системы, посвященной «инвентаризации ресурсов российских архивов, представленных в Интернете, и формированию базы данных описаний этих ресурсов». Отчасти эта задача пересекается с содержанием проекта историко-ориентированных информационных систем (в части каталога), который был реализован Центром цифровой гуманитаристики ПГНИУ (<http://digitalhistory.ru/>). К сожалению, авторы статьи не упомянули данный проект, хотя очевидно, что идея создания справочно-информационной системы по цифровой гуманитаристике (далее - СИС ЦГ (сокращение авторов)) с упором на изучение архивных информационных ресурсов является явно не оригинальной.

Авторы конкретизируют задачи разработки данной системы, указывая, что «система должна содержать информацию о созданных в мире и России объектах ЦГ в целях их повторного использования и исключения дублирования при разработке».

Рецензенту представляется, что авторам необходимо точно определить содержание многих используемых в статье терминов, начиная с термина «ресурсы российских архивов» и заканчивая термином «объект цифровой гуманитаристики».

Первый термин (в версии «архивные информационные ресурсы») уже давно введен в историографию, под ним понимается не страница, сайт, портал и т.п., а созданные на основе научно-справочного аппарата архивов информационные системы или представление в сети Интернет коллекций электронных копий архивных документов в виде тематических публикаций, виртуальных выставок и т.п.

Анализ служебных разделов порталов, сайтов, страниц в настоящее время утратил смысл, поскольку у архивов уже накоплен достаточный опыт разработки этих информационных блоков, к тому же содержание этих разделов подвержено быстрым информационным изменениям в результате организационных и иных пертурбаций, происходящих в отрасли.

Второй термин употребляется авторами в тексте то в качестве аналога понятия архивного информационного ресурса (АИР), то в качестве синонима к понятию комплекса конкретных архивных документов, что очень путает читателя.

Большое количество вопросов вызывают перечисленные возможности использования создаваемой СИС ЦГ, которая, по мысли авторов «может быть использована:

- при разработке программ НИР;
- при подготовке и экспертизе заявок на гранты, проекты, темы НИР;
- для оценки результатов НИР, а также учреждений и научных подразделений;

- для координации работ по оцифровке объектах ЦГ;
 - для информационного обеспечения научной и образовательной деятельности по применению компьютерных технологий в области гуманитарных и общественных наук».
- Обозначенные возможности применительно к АИР вызывают недоумение:
- во-первых, если речь идет об инвентаризации архивных информационных ресурсов российских архивов, то они разрабатываются, как правило, самими архивами, которые не являются научными учреждениями и не выполняют НИР (см. ФЗ № 125 Федеральный закон от 22.10.2004 №125-ФЗ «Об архивном деле в Российской Федерации»), а выполняют свои обязанности на основе Государственного Задания и определенных им объемов финансирования;
 - во-вторых, архивы хранят архивные документы, а не «объекты ЦГ». К тому же, архивные документы (за редким исключением) не имеют дубликатов. Это означает, что создание любых информационных систем (если они подразумеваются под термином «Объекты ЦГ») на основе архивных документов исключает дублирование информации, а вот повторение тематики (например – электронные копии метрических книг) не только имеет смысл, а является чрезвычайно важным и необходимым;
 - в-третьих, если объекты «цифровые», то, вероятно, нет смысла их оцифровывать, тем более, «координировать» работы по оцифровке (если под этим термином подразумеваются работы по сканированию архивных документов или созданию ИПС на основе НСА), которые осуществляются в каждом архиве в соответствии с внутренним планом и выделяемым финансированием (Методические рекомендации по электронному копированию архивных документов и управлению полученным информационным массивом. Росархив, ВНИИДАД. 2012 г.)

Высокая цель, обозначенная авторами для разрабатываемой системы - «СИС ЦГ должна обеспечивать повышение видимости, цитируемости и повторного использования создаваемых в России объектов ЦГ» - достижима только в случае, если данная СИС и включенные в нее «объекты ЦГ», будут соответствовать требованиям ст. 2 №-149-ФЗ «Об информации, информационных технологиях и о защите информации» (с изм. и доп., вступ. в силу с 01.01.2022) и положениям Закона РФ от 27.12.1991 № 2124-1 (ред. от 01.07.2021) «О средствах массовой информации», вступившими в силу с 01.08.2021, о чем, к сожалению, в статье нет ни слова.

Весьма неоднозначно применительно к архивам и архивным информационным ресурсам выглядит перечисление объектов исследования, а также определение ресурса, под которым авторы понимают «поименованные (выделенные) массивы документов (?) /данных (?), которые и являются объектами учета и поиска» и идентификатором которых «обычно выступает сетевой адрес или URL».

Большое сожаление вызывает то, что авторы очень далеки от понимания основных направлений деятельности архивной сферы Российской Федерации, в результате чего среди информационных ресурсов органов управления архивным делом федерального и регионального уровня, а также конкретных федеральных, региональных и ведомственных архивов (Табл.1) присутствует странное обобщение ИПС (?) (вместо ЦФК, путеводители, описи, указатели и т.п.), а также совершенно не понятные («глухие») ресурсы типа «хронологии» (?), «электронные издания»(?), «медиа» (?), «музейные коллекции» (?), «изображения» (?) и т.п.

В целом, статья производит странное впечатление – совершенно очевидно, что предлагаемая к разработке СИС ЦГ в таком виде вряд ли достигнет заявленной цели – «обеспечения повышения видимости, цитируемости и повторного использования создаваемых в России объектов ЦГ», в частности, явно не поможет историкам в поиске и выявлении архивной информации.

В завершение рецензент рекомендует авторам изучить предметную область архивного (а

в будущем и музейного, и библиотечного) дела, поскольку особенности этих областей диктуют специфику создаваемых на основе электронных копий архивных документов и НСА, электронных копий музейных предметов и библиотечных фондов информационных ресурсов, и соответственно, будут влиять на содержательную структуру СИС ЦГ, в которой предполагается "учитывать" и "описывать" данные ресурсы.

По мнению рецензента, данная статья не соответствует профилю журнала "Историческая информатика".

Результаты процедуры повторного рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Статья представляет собой изложение промежуточных результатов создания справочно-информационной системы по цифровой гуманитаристике (СИС ЦГ), создаваемой в рамках совместной деятельности ИНИОН РАН и Сибирского федерального университета. В ней также излагаются результаты инвентаризации ресурсов российских архивов, представленных в Интернете, на основе которой создается база данных описаний этих ресурсов.

Работа содержит описание отбора информационных объектов (архивные ресурсы) для базы данных, их классификацию, изложение исходного варианта структуры самой базы и т. д. Таким образом можно говорить о применении информационного подхода, классификационного метода и технологии баз данных.

Актуальность работы заключается в необходимости создания разного рода справочников гуманитарных ресурсов в широком смысле слова, которые могут быть использованы при планировании научно-исследовательской работы, а также для обеспечения научно-образовательной, информационной и других смежных видов деятельности.

Все сказанной об актуальности справедливо отнести и к новизне исследования, поскольку подобная справочно-информационная система в России пока не создавалась. Сравнительно новой является и постановка некоторых частных методических вопросов.

Несмотря на не очень большой объем статьи, она хорошо структурирована и состоит из более, чем 10 разделов, названия которых хорошо отражают логику исследования. После вводной части обозначается цель исследования и излагается история вопроса. Далее уточняются сфера и объекты исследования (ресурсы федеральных и государственных архивов, сервисы, инструменты, проекты и пр.). Подробно рассматривается уровень детализации ресурсов, который окончательно будет установлен в ходе эксплуатации справочно-информационной системы, исходя из нужд пользователей. Далее приводится примерный состав ресурсов на основе ключевых слов. В следующих разделах речь идет о структуре базы данных и языках индексирования. Статья завершается таблицей, отражающей общее количество выявленных информационных объектов, и общими выводами. Статья написана хорошим научным стилем, академическим языком, адекватным для тематики работы. Есть небольшие опечатки.

Библиография статьи содержит некоторое количество ключевых для темы исследования работ и ссылок на информационные ресурсы. В то же время она не представляется избыточной и вполне может быть увеличена количественно.

Статья не содержит элементов прямой дискуссии с оппонентами, но является дискуссионной и обращенной к аудитории по всем обсуждаемым вопросам, связанным с функционированием создающейся справочно-информационной системы по цифровой гуманитаристике.

Статья может вызвать интерес у читателей в силу своей актуальности. Она содержит подробное описание справочно-информационной системы цифровой гуманитаристики и информационных ресурсов российских архивов. Статья рекомендуется к публикации.