Министерство науки и высшего образования Российской Федерации
Сибирский федеральный университет

ПАРАД ЦИФРОВЫХ ГУМАНИТАРНЫХ ПРОЕКТОВ

Монография

Красноярск
СФУ
2025

УДК 009:004.9+378.147:004
ББК 71.034+32.97+74.480
П180

Авторы:
А. Е. Арефьев, Т. А. Базарова, Е. В. Баранова, А. А. Бонч-Осмоловская, Е. А. Брюханова, Т. Я. Валетов, В. А. Верещагин, И. В. Владимирова, А. Ю. Володин, Ю. О. Вронская, И. С. Гурьянов, К. И. Коваленко, П. В. Колозариди, М. А. Лаптева, Т. С. Лукашевский, В. Н. Маслов, М. А. Мельниченко, А. Д. Муждаба, Б. В. Орехов, А. Ю. Павловская, М. С. Панкова, М. Е. Проскурякова, М. Е. Романюк, Е. М. Северина, П. О. Сенотрусова, А. А. Сенюхин, Ф. Н. Толстая, Д. В. Тюрякова-Матвеева, И. В. Усалко, Г. А. Шерстнев

Рецензенты:
Аликберов Аликбер Калабекович, доктор исторических наук, директор Института востоковедения Российской академии наук;
Кончаков Роман Борисович, кандидат исторических наук, декан историко-филологического факультета Института общественных наук Российской академии народного хозяйства и государственной службы при Президенте РФ

П180 **Парад цифровых гуманитарных проектов** : монография / А. Е. Арефьев, Т. А. Базарова, Е. В. Баранова [и др.] ; отв. ред. А. Ю. Володин. – Красноярск : Сиб. федер. ун-т, 2025. – 286 с.

ISBN 978-5-7638-5182-3

Представлен обзор современных российских исследований на стыке цифровых технологий и гуманитарных наук. Рассмотрен широкий спектр проектов – от создания цифровых архивов, корпусов текстов (включающих наследие А. С. Пушкина, Л. Н. Толстого, А. П. Чехова, дневники проекта «Прожито») и коллекций историко-культурного наследия (проект «Сибириана») до разработки баз данных по материалам переписей и этноботаническим источникам, виртуального моделирования исторических локаций и применения технологий искусственного интеллекта (в проекте «*Digital* Пётр»). Особое внимание уделено методологическим вызовам, технологическим решениям и интерпретационным возможностям, которые открываются перед исследователями в эпоху цифровизации. Отражены состояние и тенденции развития цифровых гуманитарных исследований в России, объединен под одной обложкой опыт ведущих университетских и академических центров. Продемонстрировано, как современные технологии позволяют по-новому работать с историко-культурным наследием, с одной стороны, обеспечивая его сохранность и доступность, а с другой – включая его в актуальный научный и общественный контекст.

Предназначена всем, кто интересуется трансформацией гуманитарного знания в цифровую эпоху, будет интересна специалистам в области цифровых гуманитарных исследований, историкам, филологам, лингвистам, философам, социологам, культурологам, архивистам, музеологам, библиотекарям, а также студентам и аспирантам.

На обложке изображён прожектор, освещающий путь инициаторов и сотрудников цифровых гуманитарных проектов. Его образ навеян зарисовками опытов со светом и линзами в записных книжках Леонардо да Винчи. Автор обложки – дизайнер ***Илья Гурьянов***.

Электронный вариант издания см.:
http://catalog.sfu-kras.ru

УДК 009:004.9+378.147:004
ББК 71.034+32.97+74.480

© Сибирский федеральный университет, 2025
© Коллектив авторов, 2025

ISBN 978-5-7638-5182-3

Содержание

Предисловие 4

Андрей Володин
Цифровые гуманитарные проекты: прожекторы, модусы, ризома и прогноз 6

Борис Орехов
20 лет цифровых гуманитарных проектов: мотивация и риски 34

Полина Колозариди
«Пушкин цифровой»: за границей текста 41

Полина Сенотрусова, Анатолий Арефьев
«Сибириана»: зазеркалье провинциальных музеев 63

Татьяна Базарова, Мария Проскурякова
«Digital Пётр»: рукописное наследие Петра Великого и технологии искусственного интеллекта 73

Елена Северина, Марина Панкова
«Цифровой Чехов»: от текста к данным, или зачем мы делаем проект Chekhov Digital 89

Кира Коваленко, Иван Усалко
Этноботаническая база данных PhytoLex: история проекта и перспективы развития 100

Елена Баранова, Виталий Маслов, Вячеслав Верещагин
Технологические вызовы и ограничения в создании виртуальных моделей бытовых локаций Кёнигсберга-Калининграда 120

Елена Брюханова
Информационная система по материалам Первой всеобщей переписи населения Российской империи 1897 года: рефлексия проектов и ресурсов 139

Тимур Валетов
Цифровой атлас административно-территориального деления России/СССР в XX веке 157

Михаил Мельниченко, Алексей Сенюхин
Корпус дневниковых записей «Прожито» 193

Андрей Муждаба, Анастасия Павловская, Георгий Шерстнев
Цифровой архив «Прожито» 206

Ирина Владимирова, Марина Лаптева, Мария Романюк
«Заботы и забавы маленьких енисейцев»: детский издательский проект с применением цифровых технологий 221

Илья Гурьянов
Цифровая летопись городов Красноярского края 235

Анастасия Бонч-Осмоловская, Юлия Вронская, Тимофей Лукашевский, Фекла Толстая, Дарья Тюрякова-Матвеева
Слово Толстого: от раритетного 90-томника к общедоступной цифровой экосистеме. Сохранение, обогащение и популяризация толстовского наследия 248

Сведения об авторах 282

20 ЛЕТ ЦИФРОВЫХ ГУМАНИТАРНЫХ ПРОЕКТОВ: МОТИВАЦИЯ И РИСКИ

Борис Орехов
Национальный исследовательский университет «Высшая школа экономики»,
Институт русской литературы (Пушкинский дом) РАН

Хотя формально ещё не прошло 20 лет с тех пор, как я начал разрабатывать цифровые гуманитарные проекты, я хочу прибегнуть к риторическому приёму и округлить цифру, тем более что время течёт быстро и совсем скоро этот символический рубеж будет преодолён.

Опираясь на накопившийся опыт, я попробую подвести предварительные итоги почти 20-летней работы, сосредоточившись на мотивации, которая лежала в основе создания проектов, и рисках, которые этим проектам сопутствовали.

Параллельный корпус переводов «Слова о полку Игореве» запущен в феврале 2007 г.[1] Это электронный инструмент сравнения текстов, которые для его создания ещё предстояло оцифровать. Идея заключалась в том, чтобы корпус позволил с бóльшим удобством, чем без цифровой поддержки, заниматься медленным чтением переводов средневекового памятника, количество которых исчисляется многими десятками. Сама концепция такого инструмента не выглядит чрезмерно прорывной, особенно если учитывать, что цифровая форма служит только для комфорта исследователя, пользующегося традиционными методами, и не предполагает современных компьютерных подходов к обработке текста и анализу данных.

С технической точки зрения ресурс также не переусложнён. Цифровая платформа должна была предоставлять только функции выбора выводимого на экран контента и собственно электронные формы его представления. И то, и другое вписывалось в базовую

[1] Орехов Б. В. Параллельный корпус переводов «Слова о полку Игореве»: итоги и перспективы // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009. С. 462–473.

© Орехов Борис, 2025

функциональность веб-разработки с самого начала появления интернета.

Мотивация создания этого проекта состояла в том, чтобы обеспечить филолога удобным инструментом для его работы. Одновременно с этим не было акторов, которые располагали бы нужными компетенциями и поддержали бы эту инициативу. В середине 2000-х в сети было не так много специализированных ресурсов для филологической работы. Главным образом, они дублировали аналоговые носители в цифровой форме: на должном уровне развития находились электронные библиотеки, в т. ч. профессиональные (РВБ, ФЭБ). Недавним явлением стали снабжённые сложной системой поиска корпуса, обслуживающие интересы лингвистов. Ресурс, включающий тексты переводов «Слова о полку Игореве», не принадлежал ни к одному из этих классов, т. к. не предполагал поиска и морфологической разметки в отличие от корпусов (в корпус необходимо было включить переводы на десятки языков, обеспечить их все морфологическим разбором было бы невозможно, особенно во второй половине 2000-х) и не воспроизводил все особенности печатного издания в отличие от электронной библиотеки. Более того, текст в ряде случаев видоизменялся, что было недопустимо для идеологии работы с источниками, положенной в основу электронных библиотек. Изменения, разумеется, не предполагали редактирование или подобные действия, которые можно охарактеризовать как вмешательство в текст. Но для того, чтобы представить тексты в параллельном режиме, переставленные переводчиками фрагменты приходилось унифицировать (см. начало фрагмента № 8, которым открывается описание затмения). Таким образом, речь шла о создании ресурса нового типа, и реализацию такой идеи в тот момент не был готов взять на себя ни один из существующих коллективов, занимавшихся разработкой цифровых ресурсов. Сейчас можно отметить и побочную мотивацию – знакомство с малоизвестными переводами памятника и введение их в (не только научный) оборот путём оцифровки.

Однако результаты работы в этой системе ориентиров оказались неожиданными. Ресурс действительно стал востребован в аудитории филологов, но не как инструмент сравнения, а как хранилище текстов. На момент запуска Параллельного корпуса в сети уже было в разрозненном виде представлено около полутора десятков переводов памятника на современный русский язык и некоторое штучное

количество текстов на других языках. Работа над корпусом позволила оцифровать несколько десятков других переводов и тем самым дать к ним доступ широкой аудитории. Именно эта функция, по сути дублирующая функцию электронной библиотеки, и оказалась для ресурса основной. Здесь я бы отметил возможный риск, сопутствующий цифровому гуманитарному проекту: изначально спроектированные для него практики использования могут оказаться не нужны реальным пользователям. Особенно если речь идёт об узком исследовательском сегменте научного поля, таком как *translation studies*. Параллельному корпусу переводов «Слова о полку Игореве» в некотором смысле повезло, что подразумевавшаяся для него, но не востребованная функциональность заместилась практикой работы с другим классом ресурсов (электронных библиотек). Обстоятельства могли бы сложиться менее благополучно, и тогда корпус совсем не использовался бы гуманитариями (такие случаи упомянуты ниже).

Но с принятием на себя функциональности электронной библиотеки корпус принял на себя и риски, сопутствующие ресурсам такого класса.

Во-первых, его сопровождает неурегулированность ситуации с авторскими правами. Хотя за 18 лет существования корпуса не было ни одного случая конфликта с правообладателями, потенциально такие конфликты возможны. В отличие от лингвистических корпусов, которые редко показывают содержимое документов целиком и обеспечивают демонстрацию небольшого фрагмента, достаточного для добросовестного цитирования, в корпусе переводов «Слова» есть возможность ознакомиться с полным текстом произведения, что, собственно, и стало его основным назначением. Вероятно, потенциальный конфликт не поставит под угрозу существование всего ресурса, а только сделает невозможным демонстрацию на сайте защищённого авторским правом текста, но всё же к такого рода ситуациям приходится быть готовым.

Во-вторых, работа корпуса в слишком большой степени зависит от одного человека, обеспечивающего его техническую и финансовую поддержку. Хотя в период активной разработки и оцифровки (2008) проект был поддержан грантом РГНФ, что означает, что создание корпуса могло финансироваться только через организацию (в этой роли выступал Башкирский государственный педаго-

гический университет им. М. Акмуллы), после окончания грантового периода ресурс более не ассоциирован ни с одной институцией и существует исключительно благодаря усилиям создателя. После того как создатель утратит возможность поддерживать работоспособность корпуса (например, оплачивать хостинг), существование ресурса окажется под вопросом.

Фольклорный архив Башкирского государственного университета разработан в 2011 г.[1] Несмотря на название, в настоящий момент он так же, как и предыдущий описанный здесь проект, не ассоциирован с институциями и находится в зоне риска, связанного с возможностями одного человека. В рамках этого электронного издания оцифровывались материалы фольклорных экспедиций, осуществлявшихся сотрудниками и студентами Башкирского университета, начиная с первой половины XX в. (на сайте есть три текста, записанных в 1940-е). В отличие от традиционных электронных библиотек, в функциональность ресурса были встроены элементы поиска, сопроводительные геоинформационные материалы, конкорданс. Эти модули существенно расширяли исследовательские возможности, выходящие за рамки традиционной концепции дипломатического воспроизведения печатного источника.

Мотивацией этого проекта стала всеобщая цифровизация, особенно актуальная для наследия миноритарных народов. Именно цифровое пространство способно обеспечить связность информации в ситуации господства престижной культурно-языковой среды.

Тем не менее ресурс оказался ещё в худшем положении, чем Параллельный корпус переводов «Слова о полку Игореве». Фольклорный архив был полностью проигнорирован сообществом учёных, изучающих башкирскую культуру: в позднейшей книге[2] ошибочно утверждается, что на её страницах размещён первый опыт создания конкорданса на основе текстов башкирского песенного фольклора, в то время как первый такой конкорданс появился на пять лет раньше в рамках описываемого электронного издания. Такое положение вещей относится к возможным рискам цифровых проектов, кото-

[1] Галлямов А. А., Орехов Б. В. Об электронном научном издании «Фольклорный архив Башкирского государственного университета» // Вестник РГГУ. История. Филология. Культурология. Востоковедение. 2016. № 12. С. 140–149.

[2] Муратова Р. Т. Словарь-конкорданс башкирских народных песен о Родине. Уфа, 2016. С. 13.

рые в силу разных причин могут маргинализироваться и отторгаться сообществом.

Первоначально ресурс развивался в рамках инфраструктурных решений Башкирского университета, что и обусловило зафиксированный в публикации (см. выше) *URL*: lcph.bashedu.ru/index.php?go=editions. Но со временем разработчики потеряли доступ к серверным мощностям БашГУ, а выделенный под издание поддомен *lcph.bashedu.ru* был отключён. Проект был перенесён на новый *URL*: nevmenandr.net/pages/bashfolk.php. Но его поиск теперь затруднён, т. к. прямой путь из упомянутой научной публикации к действующему сервису более недоступен. Такого рода риски заставляют задуматься об аналоге *DOI* (*digital object identifier*) для цифровых проектов, который бы позволил всегда находить нужный ресурс, даже в случае его переноса между доменами.

Аналогичные проблемы сопровождали приложение ***91-й том***[1] (2017), к управлению *URL* которого разработчики на время потеряли доступ. В настоящий момент приложение функционирует и открывается по прежнему адресу, но устойчивость такой ситуации под вопросом.

Мотивация, которая стояла за созданием этого ресурса, касалась обновления практик работы с данными. Цифровая среда предоставляет возможности для рекотекстуализации привычных наборов данных, помещения их в новую «обёртку», которая позволяет отыскивать в массивах новые связи и закономерности.

Объектом рекотекстуализации в данном случае стали данные традиционного указателя имён и названий, составленного редакторами 90-томного Полного собрания сочинений и писем Л. Н. Толстого. С одной стороны, электронная версия указателя сохраняла весь поисковый потенциал бумажной копии (топоним, название произведения или имя упоминаемого лица всё так же можно было бы

[1] Orekhov B. "Volume 91": An electronic index to the complete works of Leo Tolstoy // Journal of Siberian Federal University. Humanities and social sciences. 2020. No. 12. Pp. 2 049–2 055. DOI: 10.17516/1997-1370-0703; Orekhov B. V., Fischer F. The 91st Volume – How the Digitised Index for the Collected Works of Leo Tolstoy Adds a New Angle for Research // Digital Humanities 2018. Puentes-Bridges. Book of Abstracts, 2018. Pp. 465–466; Орехов Б. В. Make index great again: электронный указатель к 90-томнику Л. Н. Толстого // Информационные технологии в гуманитарных науках. Тезисы докладов конференции. Красноярск: СФУ, 2017. С. 112–113.

найти в тексте), с другой, оказались актуализированы внутритекстовые связи, трудновосстановимые из привычной формы хранения данных: исследователь получил возможность отыскивать имена и названия, упоминаемые у Толстого на одном текстовом отрезке, находить насыщенные именами фрагменты текста, представлять онтологию упоминаемых Толстым имён и названий в виде сети.

Сходная мотивация стояла за созданием карты ***Недоимки по подушной подати с 1724 по 1740 г. в России*** (2013)[1]. Традиционное представление числовых данных, релевантных для экономической истории, оформлено в виде таблицы – в целом удобного, но не самого визуально богатого формата. Карта – во многом выигрышная альтернатива, позволяющая в буквальном смысле увидеть пространственные связи между элементами данных. Современные геоинформационные инструменты с низким порогом вхождения располагают к тому, чтобы строить такие представления данных сразу на цифровой платформе, обеспечивающей интерактивность, простоту копирования и переносимость. Кстати, и этот проект пережил цифровой переезд, так что актуальный адрес отличается от опубликованного в 2014 г.

Продолжают эту стилистическую линию «Сеть персоналий в Словаре русских писателей XVIII в.»[2] (2022) и «Таймлайн жизни русских писателей XVIII в.»[3] (2023). Оба ресурса созданы на основе данных, извлечённых из «Словаря русских писателей XVIII в.»[4]. Стремлением избежать рисков, связанных с жизненным циклом цифровой инфраструктуры, обусловлен выбор площадки для размещения этих проектов – *GitHub Pages*. Всё говорит о том, что это долговечная и востребованная разработчиками платформа. К счастью, оба проекта не требуют исполняемых программ на стороне сервера, поэтому сама возможность их функционирования на такой площадке доступна. Сеть персоналий необходима как интерактив-

[1] Корчмина Е. С., Орехов Б. В. Использование интерактивных карт при изучении истории налогов в России XVIII в. // Информационный бюллетень Ассоциации история и компьютер. 2014. № 42. С. 138–139.

[2] URL: nevmenandr.github.io/rus-dict18-persons.

[3] URL: nevmenandr.github.io/18cent-timeline.

[4] Орехов Б. 2022. Словарь русских писателей XVIII века: сеть персоналий. DOI: 10.31860/openlit-2022.1-B002. Репозиторий открытых данных по русской литературе и фольклору, V1.

ное дополнение к датасету, позволяющее визуально и аналитически оценить ценность заключённых в нём данных. Можно приобщить эту необходимость к тому, что ранее я назвал обновлением практики работы с данными. Таймлайн в меньшей степени пригоден для работы исследователя в качестве аналитического инструмента и в большей степени является популяризаторской визуализацией, представляющей историческую проекцию литературного процесса в XVIII в.

К рискам в этих случаях, как и с Фольклорным архивом Башкирского университета, можно отнести низкую вовлечённость в орбиту этих разработок профессионального сообщества. Несмотря на декларированную ассоциированность названных проектов с Лабораторией ИРЛИ РАН (Пушкинского дома), на юбилейном сайте к пятилетию Лаборатории[1] об этих проектах не упоминается. Нет ссылок на них и со страницы Лаборатории на сайте Института.

Персидский поэтический корпус[2] (2020), в отличие от корпуса переводов «Слова», в гораздо большей степени похож на «настоящий» корпус благодаря морфологической разметке и поисковой системе. Коллекция текстов классической средневековой персидской поэзии размещена на технологической платформе корпусного менеджера *tsakorpus*. Облик проекта во многом сформирован образцом поэтического корпуса в составе НКРЯ, и реализация проекта была мотивирована желанием познакомить востоковедов с цифровыми практиками, открытыми для себя русистами.

Корпус существует благодаря серверной инфраструктуре школы лингвистики НИУ ВШЭ, и его судьба и перспективы существования полностью зависят от ресурсов институции.

Таким образом, среди больших и малых цифровых проектов, реализованных за последние 20 лет, преобладает мотивация цифрового обновления практик работы с привычными данными, а в числе наиболее реальных рисков видятся те, которые делают невозможной работу ресурса на прежней площадке (на длинной временной дистанции реальность этих рисков видна особенно хорошо), и равнодушие со стороны потенциальных пользователей.

---

[1] URL: pushdigilab.tilda.ws.

[2] Орехов Б. В., Степина Д. С. Персидский поэтический корпус // Труды Института русского языка им. В. В. Виноградова. 2022. № 1. С. 65–71. DOI: 10.31912/pvrli-2022.1.7.

Мельниченко Михаил Анатольевич – кандидат исторических наук, руководитель Центра изучения эго-документов «Прожито» Европейского университета в Санкт-Петербурге.

Муждаба Андрей Дмитриевич – технический директор Центра изучения эго-документов «Прожито» Европейского университета в Санкт-Петербурге.

Орехов Борис Валерьевич – кандидат филологических наук, доцент Школы лингвистики факультета гуманитарных наук Национального исследовательского университета «Высшая школа экономики», старший научный сотрудник Института русской литературы (Пушкинский дом) РАН, главный редактор журнала «Цифровые гуманитарные исследования».

Павловская Анастасия Юрьевна – исследователь Центра изучения эго-документов «Прожито» Европейского университета в Санкт-Петербурге.

Панкова Марина Сергеевна – преподаватель кафедры лингвистики и профессиональной коммуникации Института филологии, журналистики и межкультурной коммуникации Южного федерального университета.

Проскурякова Мария Евгеньевна – кандидат исторических наук, старший научный сотрудник Отдела источниковедения Санкт-Петербургского института истории РАН.

Романюк Мария Евгеньевна – старший преподаватель кафедры информационных технологий в креативных и культурных индустриях Гуманитарного института Сибирского федерального университета.

Северина Елена Михайловна – доктор философских наук, профессор кафедры лингвистики и профессиональной коммуникации Института филологии, журналистики и межкультурной коммуникации Южного федерального университета.

Сенотрусова Полина Олеговна – кандидат исторических наук, доцент кафедры информационных технологий в креативных и культурных индустриях Гуманитарного института Сибирского федерального университета, старший научный сотрудник Лаборатории археологии Енисейской Сибири Гуманитарного института Сибирского федерального университета.

Сенюхин Алексей Андреевич – исследователь Центра изучения эго-документов «Прожито» Европейского университета в Санкт-Петербурге.

Толстая Фекла Никитична – руководитель группы *Tolstoy Digital*, сотрудник Государственного музея Л. Н. Толстого.

Тюрякова-Матвеева Дарья Владимировна – участник группы Tolstoy Digital, сотрудник Национального исследовательского университета «Высшая школа экономики».

Усалко Иван Викторович – сотрудник Европейского университета в Санкт-Петербурге.

Шерстнев Георгий Александрович – архивист Центра изучения эго-документов «Прожито» Европейского университета в Санкт-Петербурге.

Научное издание

ПАРАД ЦИФРОВЫХ ГУМАНИТАРНЫХ ПРОЕКТОВ

Монография

Ответственный редактор
Володин Андрей Юрьевич

Редактор *М. А. Лаптева*
Корректор *Л. В. Боос*
Компьютерная вёрстка *И. В. Манченковой*

Подписано в печать 30.10.2025. Печать плоская. Формат 60×84/16
Бумага офсетная. Усл. печ. л. 17,0. Тираж 500 экз. Заказ № 25454

Библиотечно-издательский комплекс
Сибирского федерального университета
660041, Красноярск, пр. Свободный, 82а
Тел. (391) 206-26-16; http://bik.sfu-kras.ru
E-mail: publishing_house@sfu-kras.ru