

ИСТОРИЯ В КОНТЕКСТЕ ЦИФРОВЫХ ГУМАНИТАРНЫХ ИССЛЕДОВАНИЙ

УДК 930.23

doi 10.17072/2219-3111-2023-4-135-147

Ссылка для цитирования: *Володин А. Ю.* Исторические исследования в контексте датаизма: методологический аспект // Вестник Пермского университета. История. 2023. № 4(63). С. 135–147.

ИСТОРИЧЕСКИЕ ИССЛЕДОВАНИЯ В КОНТЕКСТЕ ДАТАИЗМА: МЕТОДОЛОГИЧЕСКИЙ АСПЕКТ¹

А. Ю. Володин

Московский государственный университет имени М. В. Ломоносова, 119192, Россия, Москва, Ломоносовский пр., 27, корп. 4, Г-423;

Сибирский федеральный университет, 660041, Россия, Красноярск, пр-т Свободный, 82А (Пирамида), ауд. 4-60

volodin@hist.msu.ru

SPIN-код: 9627-6687

ResearcherID: J-9603-2012

Scopus Author ID: 57219052529

Рассматривается методологический аспект использования данных в историческом исследовании. В контексте датаизма представлены дискуссии об особенностях формализации сведений исторических источников в моделях данных. Сегодня в дискуссиях о принципах применения информационных технологий в исторических исследованиях можно встретить немало противоречий, связанных со сложностями взаимодействия подходов информатики, исторических и социальных наук, все еще встречаются противопоставления качественных и количественных подходов, несмотря на широкое распространение «смешанных методов» и подходов цифровых гуманитарных исследований (Digital Humanities), а также отмечаются различия задач информационных технологий как отрасли и научного сообщества. Данные становятся общим знаменателем таких дискуссий, потому что в современных исследованиях приходится опираться именно на них. Обсуждаются особенности данных в исторических исследованиях, которые позволяют абстрагироваться от исходных источников и собрать систематические формализованные наблюдения. Рассмотрены такие особенности данных, как формат, занимаемая память и пассивность. Особое внимание уделено таким свойствам научных данных, как объем, разнообразие форматов, скорость накопления, изменчивость, достоверность, визуализируемость и ценность (7Vs). Таким образом, исторические данные в семиотическом смысле выполняют три функции: называют свойства предметов реального мира (номинация), связывают названные свойства друг с другом (предикация), располагают названное в пространстве и времени (локация). Сделан вывод, что с методологической точки зрения в историческом исследовании данные – это метод абстрактного наблюдения, когда из разнообразных исторических источников последовательно и строго формализованно собираются систематические данные.

Ключевые слова: данные, датаизм, капта, историческая информатика, методология исторического исследования, цифровые гуманитарные исследования, цифровая история.

Историк будущего будет программистом,
или его не будет вовсе.

Эммануэль Ле Руа Ладюри (1929–2023)

«Датаизм провозглашает, что Вселенная состоит из потоков данных и что ценность всякого явления или сущности определяется их вкладом в обработку данных», – так Ю. Н. Харари описывает современную эпоху увлеченности «большими данными» в книге «Homo Deus: краткая история будущего» [Харари, 2016, с. 430]. «Датаизм разрушает барьер между животными и машинами и предрекает, – продолжает он, – что электронные алгоритмы в конце концов рас-

шифруют и превзойдут биохимические алгоритмы» [Там же]. Но смысл датаизма как подхода к современной цифровой реальности не столько в противопоставлении человека и машины, а в изменении ключевого процесса перевода зафиксированных сигналов реальности в мудрость. Классическая схема информационной иерархии состоит в том, что сигналы складываются в данные, из данных человек создает информацию, информация уже преобразовывается в знания, а знания позволяют достичь мудрости. «Датаизм переворачивает традиционную пирамиду обучения. До недавних пор на данные смотрели как на первое звено в длинной цепочке интеллектуальной деятельности. Человеку надо было превращать данные в информацию, информацию в знания, а знания в мудрость. Но датаисты считают, что люди больше не в состоянии справляться с огромными потоками данных, поэтому не могут превращать данные в информацию и уж тем более в знания или мудрость. Поэтому обработка данных должна быть доверена электронным алгоритмам, намного более мощным, чем человеческий мозг. На практике это означает, что датаисты скептически относятся к человеческим знаниям и мудрости и предпочитают полагаться на большие данные и компьютерные алгоритмы» [Там же, с. 430].

Сегодня в дискуссиях о принципах применения информационных технологий в исторических исследованиях можно встретить немало противоречий, связанных со сложностями взаимодействия подходов информатики, исторических и социальных наук, все еще встречаются противопоставления качественных и количественных подходов [Селунская и др., 2018], несмотря на широкое распространение «смешанных методов» (mixed methods), а также отмечаются различия задач информационных технологий как отрасли и научного сообщества [Milligan, 2022]. Данные становятся общим знаменателем таких дискуссий, потому что в современных исследованиях приходится опираться именно на них [Ahnert et al., 2023].

Например, Адам Кримбл, создатель и вдохновитель онлайн-проекта «Программирующий историк» [The Programming Historian, 2023], будто бы развивая мысль Э. Ле Руа Ладюри из эпиграфа к этой статье, в своей книге «Технологии и историк: трансформации в цифровую эпоху» указывает, что вся сфера профессиональной деятельности историков преобразилась в эпоху цифровых технологий: программы исследований, преподавание и обучение, научное общение, архивная работа – все претерпело кардинальные изменения, технологизацию, датафикацию, влияние цифровых гуманитарных исследований и алгоритмических подходов, тем не менее роль технологий и внимание к развитию технологий оказываются весьма ограниченными (если не брать в расчет оммаж Фрэнку Оусли или Роберто Бузе), такие вопросы обычно поднимаются в кругу небольших специализированных сообществ «цифровых историков» или «цифровых гуманитариев» [Crymble, 2021].

Попробуем отказаться от характерного для таких дискуссий футуристического взгляда и посмотрим на разнообразие данных в современной методологической ситуации. Меняют ли данные как основа для методологических размышлений исторические исследования сегодня?

Данные, по определению, это предоставление информации в формальном виде, пригодном для передачи, интерпретации или обработки людьми или компьютерами [ISO/IEC 2382:2015]. При этом важно учитывать, что в исторических исследованиях в большинстве случаев используются специально собранные данные, часто собранные под конкретную исследовательскую цель. Поэтому имеет смысл учесть определение гуманитарных данных, которое дал К. Шёх в статье «Большие? Умные? Чистые? Беспорядочные? Данные в гуманитарных науках»: данные в гуманитарных исследованиях – «цифровая, выборочно сконструированная, управляемая компьютером абстракция, представляющая некоторые аспекты определенного объекта гуманитарного исследования» [Schöch, 2013]. Таким образом, данные – это абстракция, которую ученый конструирует, превращая в модель как систему принципиальных свойств изучаемого объекта. Моделирование становится ключевым методом работы с такими упорядоченными абстрактными данными [Бородкин, 2016], а базовой технологией становится система управления базами данных (основополагающая роль баз данных в информатизации исторических исследований признана давно [Гарскова, 1994]).

Даже само слово «данные» создает впечатление, что данные кем-то даны. В действительности сбор данных требует много труда. Данные, особенно исторические, появляются благода-

ря скрупулезному повседневному труду. Учитывая, что историк занимается опосредованным наблюдением прошлой реальности, наши сегодняшние исторические данные напрямую зависят от того, какие исторические источники сохранились от былых эпох: без источников исторических данных не будет, разве что гипотетические.

Абстрактность данных для исторического исследования – важная характеристика, потому что она объясняет, как данные абстрагируются от исторического источника, архивного документа, археологической находки, музейного объекта, библиографической редкости. Работая с данными, мы принципиально работаем с упрощенной моделью исторической реальности. Следует это учитывать, когда с исторической точки зрения мы оцениваем самостоятельность того или иного набора данных: можно ли его использовать, не обращаясь к источникам, на основе которых он собран. Особенно явно сложность взаимоотношений модельных данных и исторического источника возникает при обсуждении так называемых «цифровых двойников», когда точность описания, представленная в данных, все равно не позволяет передать всей материальной полноты описываемого источника [Сенотрусова, Володин, 2023].

Помимо абстрактности данных, стоит учитывать их триединую характеристику: данные – это формат, то есть представление сведений в формализованном виде, пригодном для передачи и обработки в некотором информационном процессе, или спецификация структуры данных, записанных в компьютерном файле. Также данные – это память, то есть совокупность ячеек памяти, обладающих определенными свойствами. И наконец, данные, в отличие от операций (действий или процессов), пассивны, таким образом, чтобы получить результат, с данными необходимо провести какое-то исследовательское действие. Данные сами не «говорят», данные необходимо извлекать, исследовать, визуализировать, интерпретировать.

Как сформулировать профессиональное отношение к текущим процессам цифровой трансформации исторической профессии в эпоху больших языковых моделей? В дискуссиях о научных данных называют самые разные свойства, на которые необходимо обращать первостепенное внимание, чаще всего указывают на следующие: объем, разнообразие форматов, скорость накопления, изменчивость, достоверность, визуализируемость и ценность. В литературе по информационным технологиям такую классификацию для мнемонической простоты называют «великолепной семеркой V-свойств», по той причине что по-английски все перечисленные свойства начинаются на латинскую букву V (*volume, velocity, variety, veracity, variability, visualization, value*). Рассмотрим последовательно эти семь ключевых характеристик исследовательских данных.

Начнем с объема данных. Например, авторы «Исторического манифеста» верят в силу естественного интеллекта и утверждают, что необходимо быть на технологическом фронтире: «Мы живем в новую эпоху “больших данных” – от расшифровки генома человека до миллиардов слов в официальных отчетах, которые ежегодно производят правительственные учреждения. В социальных и гуманитарных науках обращение историков и социологов к “большим данным” отражает их стремление идти в ногу со временем, использовать открывающиеся возможности для решения старых вопросов и формулирования новых» [Guldi, Armitage, 2014, p. 88]. При этом нельзя не отметить, что ожидание значительных эффектов от применения больших данных в исторических исследованиях не вполне оправдалось. Возникли многочисленные проекты оцифровки больших комплексов исторических источников (например, проект «Венецианская машина времени» [Venice Time Machine Project..., 2019]), но такие коллекции оказались основой для вполне классических с методологической точки зрения исторических исследований. Вместе с ними быстро развивались новые подходы к репрезентации исторических данных, прежде всего трехмерная реконструкция и создание проектов виртуальной и дополненной реальности, но сам объем данных оказался совсем не настолько критическим, чтобы нельзя было обойтись стандартными технологиями, без привлечения подходов, характерных для использования больших данных (вроде NoSQL, MapReduce и Hadoop) [Большие данные..., 2019].

Важно отметить, что в общественном мнении бытует точка зрения, что большие оцифрованные массивы данных приведут к существенному пересмотру больших исторических процессов. Однако профессиональный взгляд на развитие исторических данных свидетельствует, что большие данные приводят скорее к нюансировке, а не к пересмотру имеющихся в научных

кругах знаний. Причем степень точности в таких случаях может показаться со стороны ничтожно малой. Тем не менее объем, конечно, может повлиять на методологию исследований, если получится историку работать не с выборкой, а с генеральной совокупностью документов, на основе которых исследователь сможет создать полную базу данных.

Но так ли важен объем данных в контексте датаизма? Чаще всего этот вопрос формулируется так: насколько большие данные действительно большие? Один из ответов предлагают авторы исследования «Исторический макроскоп»: «Большими данными для литературоведа могут быть сотни романов («великое непрочтенное»), для историка ими может оказаться весь массив судовых журналов XIX в., для археолога это может быть каждый бит данных, полученных за несколько сезонов полевых работ или камеральных исследований. Для специалиста в области компьютерных наук большие данные часто касаются таких объемов, которые сложны для обработки обычными компьютерными средствами, например, коллекция поисковой системы Google или удивительные объемы экспериментальных данных Большого адронного коллайдера ЦЕРНа» [Graham et al., 2015, p. 3]. Сегодня мы понимаем, что проблема больших данных с практической точки зрения была преувеличена, не так-то много задач, которые требуют разработки действительно специальных подходов.

Разнообразие данных, которые сегодня уже циркулируют в исторических исследованиях, наводит на мысль о необходимости дать холистическое определение историческим данным: историческими данными можно считать любые данные, которые имеют какую-либо хронологическую метку. При этом нельзя не признать, что в большинстве случаев данные рассматриваются не в абсолютном их значении, а собираются для решения конкретной исторической задачи. И часто данные просуществуют столько, сколько просуществует исследовательский проект. Хотя уже несколько десятилетий обсуждаются проекты создания репозитория исторических данных, которые бы могли постепенно накапливаться, создавая все большие возможности для сопоставительных исследований. В последнее время одной из базовых платформ для такой задачи становится проект Dataverse [The Dataverse Project, 2023], предлагающий широкие возможности для представления готовых наборов данных широкой публике. Хотя подготовка данных к такой публикации, по сути, становится отдельной работой, в которую необходимо вложить не меньше труда, чем в написание научной статьи. В качестве примеров реализации данного подхода можно назвать инициативу по сбору данных по всемирной истории [World-Historical Dataverse, 2023], коллекции данных Международного института социальной истории [IISH Data Collection, 2023] или Репозиторий открытых данных по русской литературе и фольклору [Репозиторий..., 2023].

Разнообразие данных напрямую связано с проблемами форматов сохранения данных, а также с разнообразием форм описания этих данных в онтологиях и стандартах метаданных. Это важно учитывать, потому что разнообразие данных с технической точки зрения подразделяется на три главных варианта: структурированные (таблицы, базы данных), полуструктурированные (xml, json, csv, 3D/VR-модели) и неструктурированные (тексты, изображения, аудио, видео) данные. Технологический прогресс в обработке неструктурированных данных продвигается медленно, поэтому удобный практический путь освоения данных состоит в их структурировании, поиске неочевидных структурных и нетривиальных связей. Как известно, согласно первому информационному правилу Э. Кодда, вся информация в реляционной базе данных на логическом уровне должна быть явно представлена единственным способом – значениями в таблицах [Codd, 1970]. Вслед за этим правилом можно заметить, что сегодня создание данных в истории (да и шире – в цифровых гуманитарных исследованиях) связано с табуляцией, приведением наблюдений к табличному виду; даже семантическая разметка может быть превращена в таблицу. Таким образом, таблица является фактическим стандартом представления данных.

Скорость накопления данных в исторических исследованиях напрямую связана с двумя факторами: доступностью комплексов исторических источников и их востребованностью. С исторической точки зрения, конечно, явной проблемой становится вопрос сохранности накапливаемых сегодня данных. Говоря об исторических данных, следует учитывать, что они не такие уж и большие, если мы их сравниваем с коллекциями других современных данных. Исторические данные не накапливаются поточно. Но очевидно, что сегодня мы в состоянии

накопить такие коллекции данных, которые оказываются существенно больше возможностей нашего физического восприятия. В методическом смысле работает принцип достаточности: данных должно быть достаточно, и они должны быть репрезентативными, чтобы с их помощью можно было бы решить ту или иную исследовательскую задачу. В книге Кристины Боргман «Большие данные. Малые данные. Нет данных» ясно представлено соотношение больших данных с данными малыми [Borgman, 2015]. Причем часто в исторических исследованиях складывается ситуация, когда данных просто нет, и добыть их неоткуда, разве что получить через условные прокси-показатели (самыми известными примерами в социальной истории могут служить исследования связи антропометрических данных и благосостояния или грамотности в связи с распределением объявленного возраста в определенной популяции). Принципиальное отсутствие данных так же важно учитывать в современной методологии, как и многообразие ошибок данных, которые появляются в ходе естественного накопления датасетов, перевода аналоговых данных в электронные, сбора и обработки таблиц, баз данных, цифровых коллекций. Я. Грегори утверждает, что современное поколение историков – первое, которое сталкивается со сложными и неполными цифровыми (или оцифрованными) источниками, кишащими ошибками оцифровки, распознавания и кодирования [Gregory, 2014].

При этом большая часть данных, которые используются в исторических исследованиях, являются собранными, или так называемой каптой (подробнее см. [Володин, 2019]). Капта – это образно говоря исследовательский улов, систематически собранные данные, которые отвечают на исследовательский вопрос конкретного исследования. Капта исторического исследования представляет собой выборку, случайную или естественную (когда до наших дней сохранились лишь отдельные источники); часто такая выборка определяется временными возможностями исследователя по сбору данных в архивах, библиотеках, музеях, поэтому таким наборам данных свойственны принципиальная неполнота и незавершенность. Капту легко понять через образ археологического раскопа. То, что найдено в раскопе в этом году, является последней по близости к настоящему моменту находкой, но лишь очередной на поступательном пути науки. В следующем году будут новые находки, но анализировать и интерпретировать можно только то, что есть у нас в руках сегодня. Недаром в интеллектуальном анализе данных часто используют термин «добыча данных» (data mining) – практика весьма подобная ремеслу археолога, только в сфере науки о данных.

Изменчивость данных описывает скорость их изменения с течением времени, при этом изменчивость, с точки зрения исследователя, является очевидной уязвимостью. Можно ли доверять данным, которые завтра изменятся? Можно ли использовать базу данных, которая завтра обновится? Особое звучание проблема изменчивости исторических данных получила на страницах «Журнала цифровой истории» и была названа «апдейтизмом» [Clavert, Fickers, 2022] или отношением к обновлениям в данных «ГитХаб» [GitHub, 2023]. Технологическим примером для размышлений о проблеме обновлений стал анализ опыта крупнейшего веб-сервиса для размещения и работы над проектами, преимущественно по разработке программного обеспечения. Этот веб-сервис основан на системе контроля версий, что позволяет отслеживать все обновления и изменения как в опубликованных данных, так и в программном коде. Сформулировать свое отношение к обновлению данных «Журнал цифровой истории» решил по той причине, что сам формат публикации материалов в нем оказался новым, интерактивным, позволяющим читателю во время ознакомления со статьей посмотреть и «испытать» данные, на которых построена статья. И как следствие данные, к которым обращаются читатели, должны соответствовать тому набору данных, который существовал на момент публикации статьи. Иначе обновленные данные могут вносить коррективы в прежние наблюдения, а значит уже потребуется изменять и текст. А текст статьи должен быть устойчивым к изменениям.

Понятно, что скорость доступа и интерактивные возможности современных интерфейсов позволяют вносить изменения, исправления, дополнения весьма быстро. Но с точки зрения ответственного хранения и научной точности, несомненно, необходимо выработать традицию отметки некоторых вех на пути изменения данных, чтобы можно было видеть историю изменений, как уже привычно за ней следить на проекте «ГитХаб» или в свойствах статей на страницах «Википедии».

Один из способов противостоять изменчивости данных служит технология связанных открытых данных, позволяющих добиться устойчивой связи, например, с именованными сущностями, благодаря использованию унифицированных идентификаторов ресурсов [Meghini, 2022, p. 439–446]. Создатель Всемирной паутины, автор концепции семантической сети и инициатор подхода «связанных данных» Тим Бернерс-Ли предложил так называемую 5-звездочную схему развертывания открытых данных. Самое простое – сделать свои материалы доступными онлайн в любом формате (за такие данные дается одна звезда). Если свои данные структурировать, например в табличном виде, то это уже достойно оценки в две звезды. Перевод данных в свободные (непроприетарные) форматы (как, например, XML) – это уже следующий шаг, достойный трех звезд. Самые подготовленные данные, такие, которые используют унифицированный идентификатор (так называемый URI, позволяющий указывать на любой онлайн-ресурс: документ, изображение, файл; более современный вариант идентификации – IRI или интернационализированный идентификатор ресурса). И, наконец, пяти звезд могут удостоиться открытые связанные данные (linked open data, или LOD), благодаря чему можно семантически объединять друг с другом самые разные коллекции данных [5-звездочные..., 2023, p. 439–446]. Удачным примером реализации этой технологии является проект «Викиданные» [Wikidata, 2023].

Говоря о достоверности исторических данных, уместно вспомнить рассуждение Ю. М. Лотмана о значении дешифровки для исторической профессии: «Историк обречен иметь дело с текстами. Между событием “как оно есть” и историком стоит текст, и это коренным образом меняет научную ситуацию. Текст всегда кем-то создан и представляет собой происшедшее событие, переведенное на какой-то язык. Одна и та же реальность, кодированная разными способами, даст различные – иногда противоположные – тексты. Извлечение из текста факта, из рассказа о событии – события представляет собой операцию дешифровки. Таким образом, сознавая это или нет, историк начинает с семиотических манипуляций со своим исходным материалом – текстом» [Лотман, 1994]. При этом сегодня мы смотрим на текст как на существующий в нескольких состояниях объект: рукопись, установленный печатный вариант рукописи и распознанный посимвольно электронный текст в цифровом файле инструментально дают разные возможности для исследователя. Современные достижения в области обработки естественного языка (NLP) и векторизации больших коллекций текстов позволяют применять методики анализа данных к текстам. Например, в пособии «Текст как данные» авторы отмечают, что подход к текстовым данным близок и социальным наукам, и цифровым гуманитарным исследованиям в поиске наиболее надежных методов исследований [Grimmer et al., 2022].

Систематические наблюдения вполне свойственны гуманитарным исследованиям; поиск принципов, закономерностей, шаблонов, устойчивых структур, морфологии смыслов типичны для пытливого ума исследователей гуманитариев. Важный вклад в систематизацию такого исследовательского опыта предпринял Р. Бод в монографии «Новая история гуманитарных наук: поиск принципов и закономерностей от древности до современности». Бод обоснованно утверждает, что кажущееся бесспорным противостояние между естественными науками (математическими, экспериментальными, в которых доминируют универсальные законы) и гуманитарными науками (якобы занимающимися уникальными событиями герменевтическими методами) является ошибкой, порожденной близорукой неспособностью оценить долгий путь любознательного поиска закономерностей в истории, филологии, философии, искусстве [Bod, 2016]. Такого рода систематическое наблюдение напрямую связано с данными: с практической точки зрения необходимо формализовать, нормализовать и проверить целостность, например, исторических данных, прежде чем приступать к их анализу. И в этом также следует отметить особенность исторического познания в контексте сегодняшней датафикации: большинство исторических исследований начинаются с определения доступных исторических источников, после чего начинается сбор данных, часто самый трудоемкий и длительный период в расписании ученого. При этом, как показывают социологические исследования, историки не склонны использовать «чужие» данные, кроме как в справочных целях, а сам сбор данных является важным исследовательским этапом, характеризующим профессиональное мастерство исследователя [Володин, 2017].

Достоверность приобретает новый смысл в эпоху больших языковых моделей, потому что с развитием технологии нейронных сетей возникает все больше возможностей для алгоритмической генерации текстов, изображений, аудио- и видеосообщений. Качество результатов таких упражнений искусственного интеллекта напрямую зависит от качества и разнообразия тех данных, которые были положены в основу машинного обучения нейросети [Орехов, 2023].

Визуализируемость может показаться второстепенным свойством данных, но на практике это не так, потому что возможность визуализировать закономерности данных позволяет добыть немало важных эвристических примет на этапе разведочного анализа данных. Визуализация – это прием представления данных в виде, удобном для зрительного наблюдения и анализа. Со времен «Коммерческого и политического атласа» Уильяма Плейфэра (1786) визуализация становится важным инструментом для научных наблюдений [Wainer, 2005]. Развивающаяся в XX в. изобразительная статистика – от статистического графического языка *Isotype* до современных изоцранных интерактивных визуализаций *GapMinder* – позволяет увидеть настоящую многомерность данных [Neurath, 2010; Рослинг, 2020]. Чем больше собирается коллекция данных, тем большее значение имеют возможности визуализации свойств данных, динамики изменения показателей, «выбросов» и «лакун», особенностей «поведения» переменных. О важности визуализации в работе с данными указывал и классик разведочного анализа данных Дж. Тьюки [Tukey, 1977, р. 125–204]. Известным примером типологии представления данных в визуальных формах является работа Ф. Моретти «Графики, карты, деревья: абстрактные модели в истории литературы» [Moretti, 2007].

Исторические данные часто требуют визуализации, причем форма такого представления бывает крайне разнообразной, ведь речь идет и о статистическом распределении, и о географическом картографировании, и о виртуальной реконструкции. В основе любого исторического построения будут данные из исторических источников, разница лишь в том, насколько систематически эти данные собраны и являются ли собранные данные (капта) иллюстрациями отдельных явлений или претендуют на представительность в надежде на установление закономерностей.

Э. Тафти, один из классиков современной инфографики, в книге «Взглянуть свежим взглядом: смысл, пространство, данные, истина» сравнивает визуализацию данных с лоскутным шитьем, при этом справедливо указывает на необходимость и в графическом представлении рассуждать о причинно-следственных связях и механике событий, объяснять, проводить сравнения, оценивать достоверность измерений и выводов. «Эти принципы не могут быть изменены или отменены предубеждениями, какой-либо дисциплиной или специальностью, политической и интеллектуальной модой, маркетингом или монетизацией, тем, чего желают или требуют спонсоры» [Tufte, 2020, р. 80]. Недаром именно Тафти ввел правило «соотношения данных и чернил» (*data-ink ratio*), полагая, что в визуализации данных каждая капля чернил должна отражать представляемые данные. Важность этого правила заключается в том, что пропорции в изображении сообщают многое и сразу, и, если они не точны, то могут моментально, с первого взгляда, ввести читателя в заблуждение. Сегодня принципы, возможности и ошибки визуализации данных подробно рассмотрены в специальной литературе [Cairo, 2012, 2016, 2023; Lima, 2011; Яу, 2013].

Как показывают современные исследования, научная визуализация исторических данных требует исключительной точности, потому что информация, представленная в пространстве, воспринимается зрителем многоаспектно. Образ представляет объект во всей полноте измерений, пропорций и особенностей. О. Эйде рассмотрел на примере картографирования исторических источников проблему взаимодействия нарратива и координатной сетки карты. Он обращает внимание на то, что ключевая разница состоит в том, что изображения представляют объекты в пространстве, а текст – действия во времени. Легко найти свидетельства исторических источников, которые трудно или просто невозможно отобразить на карте. Во многих повествовательных источниках еще со времен Геродота описания слабо поддаются картографированию, а мы должны выбирать между точностью, взглядом с высоты птичьего полета карты и неоднозначностью текстов исторических источников [Эйде, 2017].

Ценность данных – важное свойство, но весьма относительное, потому что оценка ценности сильно зависит от целей генерации, рецепции и обработки данных. Недаром о ценности данных принято судить в контексте теории принятия решений. В информатике ценность данных не

является предметом изучения, так как цель, определяющая ценность, предполагается заданной извне. Поэтому посмотрим на ценность данных семиотически, через призму трехчастной схемы: прагматического, синтаксического и семантического аспектов данных. Такой подход был воспринят и творчески осмыслен в отечественном источниковедении И. Д. Ковальченко в классическом труде «Методы исторического исследования» [Ковальченко, 2003, с. 121–124].

Прагматический аспект играет важную роль в оценке исторических источников, особенно когда необходимо вынести вердикт о том, чтобы их «хранить вечно». В прагматическом смысле много сиюминутности, характерной как для контекста появления источника на свет, так и для ситуации оценки ценности источника. С историческими данными оценка ценности также часто идет по линии оценки потенциальной востребованности конкретного набора данных; на основании такой гипотезы принимается решение, необходима ли публикация данных в репозитории данных или просто на персональном ресурсе историка. При этом чаще всего препятствием для публикации становится сложность качественного документирования данных.

Синтаксический аспект напрямую связан с табличными характеристиками данных. Не все исторические источники легко поддаются формализации, лучше всего для этого подходят массовые источники, имеющие устойчивый формуляр. Тем не менее вполне возможно создать даталогическую модель и для сложно синтаксически устроенных сведений источников.

В семантическом аспекте данных важно отметить соотношение явной и скрытой информации, тем более когда наборы данных позволяют по-разному сочетать отдельные показатели, предполагая относительную автономность каждого конкретного наблюдения. «В семантическом плане чрезвычайно важен вопрос о соотношении в информации, с одной стороны, той информации, извлекалась осознанно, очевидна субъекту и может быть использована им и, с другой стороны, той, которая извлечена и зафиксирована произвольно и потому неочевидна, непосредственно не выражена для субъекта» [Там же, с. 123].

Работая с историческими источниками, накапливая собственную коллекцию (капту), создавая базу данных систематических наблюдений, историк должен обязательно учитывать принципиальные свойства семантической модели изучаемого явления. Сегодня в информационных технологиях накоплен большой опыт семантического моделирования данных, которые логично использовать в исследовательской работе историка. П. Алексопулос предлагает сосредоточиться на следующих свойствах: семантическая точность, полнота, последовательность, лаконичность, релевантность, понятность, надежность, доступность, универсальность и производительность [Alexopoulos, 2020].

Как оценить ценность данных для научного исследования? Можно сформулировать несколько подходов к такой оценке. Субъективный подход прост: каждый исследователь, собирая данные, делает прогноз, будет ли сам использовать эти данные в своем исследовании. Какие-то данные оказываются особо ценными, какие-то собираются впрок, а какие-то оказываются невостребованными. Коллективный подход к оценке данных состоит в том, есть ли надежда на повторное или, возможно, многократное использование определенного набора данных, или есть шанс, что данные будут пополняться. В таком случае публикация данных для постоянного доступа на серверах какого-либо репозитория кажется оправданной. Экономический подход (его иногда называют объективным) отвечает на вопрос: а есть ли экономический смысл в использовании данных и их долгосрочном хранении? На этот вопрос пока сложно ответить определенно. Дж. Ланье в книге «Кому принадлежит будущее?» (вполне в духе датаизма изучая ядовитую концентрацию денег и власти в цифровых сетях) отмечает, что данные уже принципиально чрезвычайно ценны, а вскоре их реальная цена вырастет настолько, что позволит человеку жить на эти средства, пока его данные используются. Но пока ученые – собиратели данных не могут выйти на открытый рынок, сегодня можно лишь ожидать, что кто-то возьмет наши данные, поэтому мы и не знаем истинную ценность данных, которая гипотетически велика [Lanier, 2014].

Таким образом, исторические данные в семиотическом смысле выполняют три функции: называют свойства предметов реального мира (номинация), связывают названные свойства друг с другом (предикация), располагают названное в пространстве и времени (локация). С методологической точки зрения в историческом исследовании данные – это метод абстрактного

наблюдения, когда из разнообразных исторических источников последовательно и строго формализованно собираются систематические данные. Большинство исторических исследований опирается на собранные комплексы данных, ответственность за качество сбора, точность формализации и успешное сохранение явных и скрытых связей, имеющихся в исторических источниках, полностью остается на совести автора-исследователя. Данные становятся моделью, которая с разной степенью полноты отражает сведения о реальных объектах прошлого. По сути, мы не уходим от знаменитой максимы Ш.-В. Ланглуа и Ш. Сеньбоса «история пишется по документам», но документы все чаще превращаются в наборы данных: текстов, семантически размеченных корпусов, сложносочиненных коллекций изображений, архивов кинофотодокументов, трехмерных реконструкций.

Конечно, в современном инструментарии ученого-историка есть немало методов, сегодня часто обобщаемых названием «цифровые гуманитарные исследования» [Alvarado, 2022; Debates in the Digital Humanities, 2023]. Можно сказать, что сегодня сложился узус (чтобы не называть перечень методов каноном), общепринятый набор подходов к анализу данных. Современный взгляд на перечень этих подходов представлен в коллективной монографии «Цифровые гуманитарные исследования» [Цифровые..., 2023]. Большая часть методов вполне доступна для применения в исторических исследованиях – это и базы данных, и географические информационные системы, анализ текстов, сетевой анализ, виртуальные реконструкции. Несмотря на установленные дисциплинарные различия в цифровых подходах истории и цифровой гуманитаристике [Робертсон, 2016], данные становятся общей основой, позволяющей расширить научный инструментарий историка.

Размышляя о будущем историка-программиста, нельзя не признать, что все мы сегодня в той или иной степени зависим от компьютерных программ, которыми пользуемся [Manovich, 2013]. Но не стоит и забывать про афоризм, что в любой программе всегда есть еще одна ошибка. Как писал один из основоположников теории искусственного интеллекта Н. Винер: «Отдайте же человеку – человеческое, а вычислительной машине – машинное. В этом и должна, по-видимому, заключаться разумная линия поведения при организации совместных действий людей и машин. Линия эта в равной мере далека и от устремлений машинопочклонников, и от воззрений тех, кто во всяком использовании механических помощников в умственной деятельности усматривает кощунство и принижение человека» [Винер, 1966, с. 81–82]. В контексте датаизма ценность историка определяется как раз вкладом в поиск, сбор и исследование сложных исторических данных. Вычислительные машины – отличные помощники в этом деле, но любопытство, понимание контекста, критическое мышление, этическое суждение и главное – формулировка исследовательского вопроса – остаются на стороне историка.

Примечания

¹ В статье использованы результаты, полученные в ходе выполнения стратегического проекта «Институт цифровых гуманитарных исследований» при поддержке Программы развития ФГАОУ ВО «Сибирского федерального университета» на 2021–2030 годы в рамках реализации программы стратегического академического лидерства «Приоритет – 2030».

Библиографический список

- Большие данные в социальных и гуманитарных науках: сб. обзоров и рефератов / РАН. ИНИОН. М., 2019. 193 с.
- Бородкин Л.И. Моделирование исторических процессов: от реконструкции реальности к анализу альтернатив. СПб.: Алетейя, 2016. 304 с.
- Винер Н. Творец и робот: обсуждение некоторых проблем, в которых кибернетика сталкивается с религией. М.: Прогресс, 1966. 104 с.
- Володин А.Ю. Цифровые практики ученых-гуманитариев: результаты онлайн-исследования [Электронный ресурс] // История. 2017. Т. 8, вып. 7(61). URL: <https://history.jes.su/s207987840001967-2-1/> (дата обращения: 01.01.2023).
- Володин А.Ю. Между data и carta: проблемы датафикации исторических исследований // Вестник Перм. ун-та. История. 2019. № 3. С. 137–145.

- Гарскова И.М. Базы и банки данных в исторических исследованиях. Гёттинген: Konrad Pachticke Max-Planck-Institut für Geschichte, 1994. 215 с.
- Ковальченко И.Д. Методы исторического исследования. 2-е изд., доп. М.: Наука, 2003. 486 с.
- Лотман Ю.М. Изъявление Господне или азартная игра? (Закономерное и случайное в историческом процессе) // Ю.М. Лотман и тартуско-московская семиотическая школа. М., 1994. С. 353–354.
- Орехов Б.В. Машинный диалект: границы рефлексии [Электронный ресурс] // Социодиггер. 2023. Т. 4, вып. 5-6 (26). URL: <https://sociodigger.ru/articles/articles-page/mashinnyi-dialekt-granicy-refleksii> (дата обращения: 01.11.2023).
- Репозиторий открытых данных по русской литературе и фольклору [Электронный ресурс]. 2023. URL: <http://dataverse.pushdom.ru> (дата обращения: 01.11.2023).
- Робертсон С. Различия между цифровыми гуманитарными науками и цифровой историей [Электронный ресурс] // История. 2016. Т. 7, вып. 7(51). URL: <https://history.jes.su/s207987840001648-1-1/> (дата обращения: 01.11.2023).
- Рослинг Х. Фактологичность. Десять причин наших заблуждений о мире – и почему все не так плохо, как кажется. М.: Corrus, 2020. 400 с.
- Селунская Н.Б., Петрова О.С., Карагодин А.В. Измерение прошлого. СПб.: Алетейя, 2018. 242 с.
- Сенотрусова П.О., Володин А.Ю. Цифровые двойники в современной археологической практике: поиск новых возможностей [Электронный ресурс] // История. 2023. Т. 14, вып. 6 (128). URL: <https://history.jes.su/s207987840027167-2-1/> (дата обращения: 01.11.2023).
- Харари Ю.Н. Homo Deus: краткая история будущего. М.: Синдбад, 2018. 496 с.
- Цифровые гуманитарные исследования: монография / А.Б. Антопольский, А.А. Бонч-Осмоловская, Л.И. Бородкин [и др.]. Красноярск: Сиб. федер. ун-т, 2023. 272 с.
- Эйде О. Вербальное выражение географической информации [Электронный ресурс] // История. 2017. Т. 8, вып. 7(61). URL: <https://history.jes.su/s207987840001948-1-1/> (дата обращения: 01.11.2023).
- Яу Н. Искусство визуализации. Как представить сложную информацию простыми способами. М.: МИФ, 2013. 352 с.
- 5-звездочные открытые данные [Электронный ресурс]. 2023. URL: <https://5stardata.info/ru/> (дата обращения: 01.11.2023).
- ISO/IEC 2382:2015. Межгосударственный стандарт. Информационные технологии. Словарь (ГОСТ 33707–2016) [Электронный ресурс]. URL: <https://docs.cntd.ru/document/1200139532> (дата обращения: 01.11.2023).
- Ahnert R., Griffin E., Ridge M., Tolfo G. Collaborative Historical Research in the Age of Big Data: Lessons from an Interdisciplinary Project (Elements in Historical Theory and Practice). Cambridge: Cambridge University Press, 2023. 80 p.
- Alexopoulos P. Semantic Modeling for Data: Avoiding Pitfalls and Breaking Dilemmas. O'Reilly Media, 2020. 328 p.
- Alvarado R. Datawork and the Future of Digital Humanities // The Bloomsbury Handbook to the Digital Humanities (Bloomsbury Handbooks). Bloomsbury Academic, 2022. P. 361–372.
- Bod R. A New History of the Humanities: The Search for Principles and Patterns from Antiquity to the Present. Oxford University Press, 2016. 402 p.
- Borgman C.L. Big Data, Little Data, No Data: Scholarship in the Networked World. MIT Press, 2015.
- Cairo A. Functional Art, The: An introduction to information graphics and visualization. New Riders, 2012. 384 p.
- Cairo A. The Art of Insight: How Great Visualization Designers Think. Wiley, 2023. 320 p.
- Cairo A. Truthful Art, The: Data, Charts, and Maps for Communication. New Riders, 2016. 400 p.
- Clavert F., Fickers A. Publishing Digital History Scholarship in the Era of Updatism // Journal of Digital History. 2022. No. 2(1). URL: <https://doi.org/10.1515/JDH-2022-0003?locatt=label:JDHFULL> (accessed: 01.11.2023).
- Codd E.F. A Relational Model of Data for Large Shared Data Banks // Communications of the Association for Computing Machinery. 1970. Vol. 13, no. 6. P. 377–387.
- Crymble A. Technology and the Historian. Transformations in the Digital Age. Illinois: University of Illinois, 2021. 245 p.

- Debates in the Digital Humanities / M.K. Gold, L.F. Klein (eds.). 2023. University of Minnesota Press, 2023. 520 p. URL: <https://dhdebates.gc.cuny.edu/read/debates-in-the-digital-humanities-2023/> (accessed: 01.11.2023).
- GitHub. 2023. URL: <https://github.com> (accessed: 01.11.2023).
- Graham S., Milligan I., Weingart S. Exploring Big Historical Data: The Historian's Macroscope. Icp, 2015. 306 p.
- Gregory I. Challenges and Opportunities for Digital History // *Frontiers in Digital Humanities*. 2014. Vol. 1, no. 1.
- Grimmer J., Roberts M.E., Stewart B.M. Text as Data: A New Framework for Machine Learning and the Social Sciences. Princeton University Press, 2022. 360 p.
- Guldi J., Armitage D. The History Manifesto. Cambridge University Press, 2014. 175 p. URL: <https://www.cambridge.org/core/publications/open-access/the-history-manifesto> (accessed: 01.11.2023).
- IISH Data Collection. 2023. URL: <https://datasets.iisg.amsterdam> (accessed: 01.11.2023).
- Lanier J. Who Owns the Future? Simon & Schuster, 2014. 448 p.
- Lima M. Visual Complexity: Mapping Patterns of Information. Princeton Architectural Press, 2011. 272 p.
- Manovich L. Software Takes Command. Bloomsbury Academic, 2013. 376 p.
- Meghini C. Linked Open Data & Metadata // *Handbook of Digital Public History* / ed. by S. Noiret, M. Tebeau, G. Zaagsma. Berlin: De Gruyter, 2022. P. 439–446.
- Milligan I. The Transformation of Historical Research in the Digital Age (Elements in Historical Theory and Practice). Cambridge: Cambridge University Press, 2022. 80 p.
- Moretti F. Graphs, Maps, Trees: Abstract Models for Literary History. Verso Books, 2007. 128 p.
- Neurath O. From Hieroglyphics to Isotype: A Visual Autobiography. Hyphen Press, 2010. 224 p.
- Schöch C. Big? Smart? Clean? Messy? Data in the Humanities // *Journal of Digital Humanities*. 2013. No. 2 (3). P. 2–13.
- The Dataverse Project. 2023. URL: <https://dataverse.org/> (accessed: 01.11.2023).
- The Programming Historian. 2023. URL: <https://programminghistorian.org/> (accessed: 01.11.2023).
- Tufte E.R. Seeing with Fresh Eyes: Meaning, Space, Data, Truth. Graphics Press, 2020. 176 p.
- Tukey J.W. Exploratory Data Analysis. Addison-Wesley, 1977. 506 p.
- Venice Time Machine Project – Current State of Affairs (19.09.2019). URL: <https://www.timemachine.eu/venice-time-machine-project-current-state-of-affairs> (accessed: 01.11.2023).
- Wainer H. Graphic Discovery: A Trout in the Milk and Other Visual Adventures. Princeton University Press, 2005. 192 p.
- Wikidata. 2023. URL: <https://wikidata.org/> (accessed: 01.11.2023).
- World-Historical Dataverse. 2023. URL: <https://dataverse.harvard.edu/dataverse/worldhistorical> (accessed: 01.11.2023).

Дата поступления рукописи в редакцию 01.11.2023

HISTORICAL RESEARCH IN THE CONTEXT OF DATAISM: METHODOLOGICAL ASPECT

A. Yu. Volodin

Moscow State University, Lomonosovskiy ave., 27-4, 119192, Moscow, Russia
Siberian Federal University, Svobodny ave., 82A, 660041, Krasnoyarsk, Russia
volodin@hist.msu.ru
SPIN: 9627-6687
ResearcherID: J-9603-2012
Scopus Author ID: 57219052529

The paper examines the methodological aspect of using data in historical research. In the context of dataism, discussions are presented about the features of formalizing information from historical sources in data models. Today, in debates about the principles of using information technologies in historical research, one can encounter many contradictions related to the complexities of interaction between the approaches of computer, historical and social sciences. There are still oppositions between qualitative and quantitative approaches, despite the widespread use of “mixed methods” and Digital Humanities research approaches, as well as the differences between the goals of the information technology industry and academia. Data becomes the commonplace of such discussions because actual

research relays on it. The paper examines the features of data in historical research, which make it possible to abstract from the original historical sources and collect systematic, formalized observations. Such data features as format, memory and passivity are considered. Attention is paid to such properties of research data as volume, velocity, variety, veracity, variability, visualization, value (7Vs). It argues that historical data, in a semiotic sense, performs three functions: naming the properties of objects in the real world (nomination), connecting the named properties (predication), and locating the named objects in space and time (location). From a methodological perspective, data in historical research is viewed as a method of abstract observation. It involves systematically and strictly formalizing observations from various historical sources. Despite the established disciplinary differences in digital approaches within history and digital humanities, data becomes a common basis that expands the researcher's toolbox.

Key words: data, dataism, capta, historical information science, historical research methodology, digital humanities, digital history.

Acknowledgments

¹ The reported study is funded by the Development Program of the Federal State Autonomous Educational Institution of Higher Education "Siberian Federal University" for 2021–2030 as part of the implementation of the strategic academic leadership program "Priority-2030", the strategic project "Institute of Digital Humanities Research".

References

- 5-zvezdochnye otkrytye dannye (2023) [5-star open data], available at: <https://5stardata.info/ru/> (accessed 01.11.2023).
- Ahnert, R., Griffin, E., Ridge, M. & G. Tolfo (2023), *Collaborative Historical Research in the Age of Big Data: Lessons from an Interdisciplinary Project (Elements in Historical Theory and Practice)*, Cambridge University Press, Cambridge, UK, 80 p.
- Alexopoulos, P. (2020), *Semantic Modeling for Data: Avoiding Pitfalls and Breaking Dilemmas*, O'Reilly Media, Sebastopol, USA, 328 p.
- Alvarado, R. (2022), "Datawork and the Future of Digital Humanities", in *The Bloomsbury Handbook to the Digital Humanities*, Bloomsbury Academic, London, UK, pp. 361–372.
- Bod, R. (2016), *A New History of the Humanities: The Search for Principles and Patterns from Antiquity to the Present*, Oxford University Press, Oxford, UK, 402 p.
- Bol'shie dannye v sotsial'nykh i gumanitarnykh naukakh* (2019) [Big data in social and human sciences], INION, Moscow, Russia, 193 p.
- Borgman, C.L. (2015), *Big Data, Little Data, No Data: Scholarship in the Networked World*, MIT Press, Cambridge, USA, 416 p.
- Borodkin, L.I. (2016), *Modelirovanie istoricheskikh protsessov: ot rekonstruktsii real'nosti k analizu al'ternativ* [Modeling historical processes: from reconstruction of reality to analysis of alternatives], Aletya, St. Petersburg, Russia, 304 p.
- Cairo, A. (2012), *The Functional Art: An introduction to information graphics and visualization*, New Riders, Hoboken, USA, 384 p.
- Cairo, A. (2016), *The Truthful Art: Data, Charts, and Maps for Communication*, New Riders, Hoboken, USA, 400 p.
- Cairo, A. (2023), *The Art of Insight: How Great Visualization Designers Think*, Wiley, Hoboken, USA, 320 p.
- Clavert, F. & A. Fickers (2022), "Publishing digital history scholarship in the era of updatism", *Journal of Digital History*, № 2(1), available at: <https://doi.org/10.1515/JDH-2022-0003?locatt=label:JDHFULL> (accessed 01.11.2023).
- Codd, E.F. (1970), "A Relational Model of Data for Large Shared Data Banks", *Communications of the Association for Computing Machinery: journal*, vol. 13, № 6, pp. 377–387.
- Crymble, A. (2021), *Technology and the Historian. Transformations in the Digital Age*, University of Illinois, Illinois, USA, 245 p.
- Eyde, O. (2017), "Verbal expression of geographical information", *Elektronnyy nauchno-obrazovatel'nyy zhurnal «Istoriya»*, vol. 8, № 7 (61), available at: <https://history.jes.su/s207987840001948-1-1/> (accessed 01.11.2023).
- Garskova, I.M. (1994), *Bazy i banki dannykh v istoricheskikh issledovaniyakh* [Databases and databanks in historical research], Konrad Pachnicke Max-Planck-Institut für Geschichte, Gettingen, Germany, 215 p.
- GitHub* (2023), available at: <https://github.com/>
- Gold, M.K. & L.F. Klein (eds.) (2023), *Debates in the Digital Humanities*, University of Minnesota Press, Minneapolis, USA, 520 p., available at: <https://dhdebates.gc.cuny.edu/read/debates-in-the-digital-humanities-2023/> (accessed 01.11.2023).
- Graham, S., Milligan, I. & S. Weingart (2015), *Exploring Big Historical Data: The Historian's Macroscope*, Imperial College Press, London, UK, 306 p.
- Gregory, I. (2014), "Challenges and opportunities for digital history", *Frontiers in Digital Humanities*, № 1:1.
- Grimmer, J., Roberts, M.E. & B.M. Stewart (2022), *Text as Data: A New Framework for Machine Learning and the Social Sciences*, Princeton University Press, Princeton, USA, 360 p.

- Guldi, J. & D. Armitage (2014), *The History Manifesto*, Cambridge University Press, Cambridge, UK, 175 p., available at: <https://www.cambridge.org/core/publications/open-access/the-history-manifesto> (accessed 01.11.2023).
- Harari, Yu.N. (2018), *Homo Deus: kratkaya istoriya budushchego* [Homo Deus: A Brief History of Tomorrow], Sindbad, Moscow, Russia, 496 p.
- IISH Data Collection (2023), available at: <https://datasets.iisg.amsterdam/> (accessed 01.11.2023).
- ISO/IEC 2382:2015. *Mezhhgosudarstvennyy standart. Informatsionnye tekhnologii. Slovar'* (n.d.) [Interstate Standard. Information Technology. Dictionary] (GOST 33707-2016), available at: <https://docs.cntd.ru/document/1200139532> (accessed 01.11.2023).
- Koval'chenko, I.D. (2003), *Metody istoricheskogo issledovaniya* [Methods of historical research], 2nd ed., Nauka, Moscow, Russia, 486 p.
- Lanier, J. (2014), *Who Owns the Future?*, Simon & Schuster, New York, USA, 448 p.
- Lima, M. (2011), *Visual Complexity: Mapping Patterns of Information*, Princeton Architectural Press, New York, USA, 272 p.
- Lotman, Yu.M. (1994), "A declaration of the Lord or a game of chance? (Natural and random in the historical process)", in *Yu.M. Lotman i tartusko-moskovskaya semioticheskaya shkola* [Yu.M. Lotman and the Tartu-Moscow semiotic school], Gnozis, Moscow, Russia, pp. 353–354.
- Manovich, L. (2013), *Software Takes Command*, Bloomsbury Academic, London, UK, 376 p.
- Meghini, C. (2022), "Linked Open Data & Metadata", in Noiret, S., Tebeau, M. & G. Zaagsma (eds.), *Handbook of Digital Public History*, De Gruyter, Berlin, Germany, pp. 439–446.
- Milligan, I. (2022), *The Transformation of Historical Research in the Digital Age (Elements in Historical Theory and Practice)*, Cambridge University Press, Cambridge, UK, 80 p.
- Moretti, F. (2007), *Graphs, Maps, Trees: Abstract Models for Literary History*, Verso Books, London; New York, UK; USA, 128 p.
- Neurath, O. (2010), *From Hieroglyphics to Isotype: A Visual Autobiography*, Hyphen Press, London, UK, 224 p.
- Orekhov, B.V. (2023), "Machine dialect: boundaries of reflection", *Sotsiodigger*, vol. 4, № 5–6 (26), available at: <https://sociodigger.ru/articles/articles-page/mashinnyi-dialekt-granicy-refleksii> (accessed 01.11.2023).
- Repozitoriy otkrytykh dannykh po russkoy literature i fol'kloru* (2023) [Repository of open data on Russian literature and folklore], available at: <http://dataverse.pushdom.ru/> (accessed 01.11.2023).
- Robertson, S. (2016), "The Differences between Digital History and Digital Humanities", *Elektronnyy nauchno-obrazovatel'nyy zhurnal «Istoriya»*, vol. 7, № 7 (51).
- Rosling, Kh. (2020), *Faktologichnost'. Desyat' prichin nashikh zabluzhdeniy o mire – i pochemu vse ne tak plokho, kak kazhetsya* [Factfulness. Ten Reasons We're Wrong About The World – And Why Things Are Better Than You Think], Corpus, Moscow, Russia, 400 p.
- Schöch, C. (2013), "Big? Smart? Clean? Messy? Data in the Humanities", *Journal of Digital Humanities*, № 2(3), pp. 2–13.
- Selunskaya, N.B., Petrova, O.S. & A.V. Karagodin (2018), *Izmerenie proshlogo* [Measuring the Past], Aleteya, St. Petersburg, Russia, 242 p.
- Senotrusova, P.O. & A.Yu. Volodin (2023), "Digital twins in modern archaeological practice: the search for new opportunities", *Elektronnyy nauchno-obrazovatel'nyy zhurnal «Istoriya»*, vol. 14, № 6 (128).
- The Dataverse Project* (2023), available at: <https://dataverse.org/> (accessed 01.11.2023).
- The Programming Historian* (2023), available at: <https://programminghistorian.org/> (accessed 01.11.2023).
- Tsifrovye gumanitarnye issledovaniya* [Digital Humanities], Sib. feder. un-t, Krasnoyarsk, Russia, 272 p.
- Tufte, E.R. (2020), *Seeing with Fresh Eyes: Meaning, Space, Data, Truth*, Graphics Press, Cheshire, USA, 176 p.
- Tukey, J.W. (1977), *Exploratory Data Analysis*, Addison-Wesley, Boston, USA, 506 p.
- Venice time machine project – current state of affairs* (19.09.2019), available at: <https://www.timemachine.eu/venice-time-machine-project-current-state-of-affairs/> (accessed 01.11.2023).
- Viner, N. (1966), *Tvoretz i robot: Obsuzhdenie nekotorykh problem, v kotorykh kibernetika stalkivaetsya s religiy* [Creator and robot. Discussing some of the problems in which cybernetics collides with religion], Progress, Moscow, Russia, 104 p.
- Volodin, A. Yu. (2017), "Digital practices of humanities researchers: results of online survey", *Elektronnyy nauchno-obrazovatel'nyy zhurnal «Istoriya»*, vol. 8, № 7 (61).
- Volodin, A.Yu. (2019), "Between data and capta: problems of datafication of historical research", *Vestnik Permskogo universiteta. Seriya Istoriya*, № 3, pp. 137–145.
- Wainer, H. (2005), *Graphic Discovery: A Trout in the Milk and Other Visual Adventures*, Princeton University Press, Princeton, USA, 192 p.
- Wikidata (2023), available at: <https://wikidata.org/> (accessed 01.11.2023).
- World-Historical Dataverse* (2023), available at: <https://dataverse.harvard.edu/dataverse/worldhistorical> (accessed 01.11.2023).
- Yau, N. (2013), *Iskusstvo vizualizatsii. Kak predstavit' slozhnyuyu informatsiyu prostymi sposobami* [Visualize This: The Flowing Data Guide to Design, Visualization, and Statistics], MIF, Moscow, Russia, 352 p.