

РЕЦЕНЗИИ И ОБЗОРЫ

BOOK REVIEWS

РЕЦЕНЗИЯ НА КНИГУ: MANNING PATRICK. BIG DATA IN HISTORY. PALGRAVE PIVOT, 2013

Сегодня понятие «больших данных» завоевывает внимание ученых из самых разных дисциплин. Благодаря популярной книге В. Майер-Шенберга и К. Кукьера проблема развития «больших данных» приобрела широкое общественное звучание¹. Историки не остались в стороне, в нашумевшем «Историческом манифесте» четвертая глава называется «Большие вопросы, большие данные»². Устойчивый интерес к данной проблематике существует, например, в экономической истории³. Можно сказать, что «большие данные» вошли в моду⁴. Тем интереснее рассуждения и прогнозы Патрика Мэннинга, который к большим данным шел долгим путем проб и ошибок. Мэннинг — известный исследователь колониализма и миграций в глобальном масштабе, стал пионером использования больших баз данных в сопоставительных международных исследованиях⁵.

Под «большими данными» сегодня понимают подходы, инструменты и методы обработки структурированных, неструктурированных, неопределенно структурированных данных огромных объемов и значительного многообразия с целью получения воспринимаемых человеком результатов, эффективных в условиях непрерывного притока данных, распределенных по многочисленным узлам вычислительной сети. Уже и в инструментарий историков входят специальные языки программирования (например, R), пригодные для статистической обработки «больших данных»⁶. Все чаще встречаются рассуждения о том, что «большие данные» касаются не столько собственно объема данных, сколько их разнообразия и разнородности⁷, хотя устоявшимся отличием «больших» и «малых» данных принято считать необходимость машинной обработки для «больших данных» и возможность ручной обработки «малых данных».

Работа П. Мэннинга обобщает англоязычную литературу по применению «больших данных» в исторических исследованиях начиная с первых опытов создания баз данных до наших дней. Основные надежды Мэннинга связаны с вызовами по изучению «глобального прошлого», которые стоят перед историками. В качестве главного примера для собственных рассуждений Мэннинг использует опыт проекта сотрудничества в области накопления и анализа исторической информации (CHIA —

¹ Майер-Шенберг В., Кукьер К. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим. М., 2014.

² Guldi J., Armitage D. The History Manifesto. Cambridge University Press, 2014. DOI: 10.1017/9781139923880

³ Large databases in economic history: research methods and case studies/ed. by M. Casson and N. Hashimzade. London; New York, 2013.

⁴ "Big Data": An Opportunity for Historians? (March 2012). URL: <http://www.historians.org/publications-and-directories/perspectives-on-history/march-2012/big-data-an-opportunity-for-historians>

⁵ Среди самых известных его монографий можно назвать: Manning P. Slavery, Colonialism and Economic Growth in Dahomey, 1640–1960. Cambridge University Press, 2004; Manning P. Navigating World History: Historians Create a Global Past. Palgrave Macmillan, 2003; Manning P. Migration in World History. 2nd ed. Routledge, 2012.

⁶ См., например: Mullen L. "Digital History Methods in R" [Electronic Resource]. URL: <http://lincolnmullen.com/projects/dh-r/>

⁷ Plantin J.-C., Lagoze C., Edwards P. N., Sandvig C. Big data is not about size — When data transform scholarship [Electronic Resource]. URL: <https://www.academia.edu/11959223>

Collaborative for Historical Information and Analysis)⁸. Автор выделяет несколько последовательных целей, которые стоят перед исследователями «больших данных»: 1) сбор и описание исторических данных; 2) создание детального исторического архива данных; 3) анализ и визуализация исторических данных во всемирном масштабе. При этом одной из принципиальных задач для «больших данных» может стать глобальное сравнение исторической динамики.

Схематично основные этапы работы с «большими данными» и связанные с ними миссии ученых-историков Мэннинг изобразил на рисунке 1.1 «Функции и возможности ресурса исторических цифровых данных» (с. 6), привожу в переводе на русский язык.

Работа с «большими данными» требует сотрудничества различных специалистов и состоит из нескольких принципиальных уровней: 1) получение данных (в том числе и с помощью популярного сетевого краудсорсинга⁹); 2) документирование данных (в том числе метаописание, курирование и гармонизация); 3) обработка данных (включая агрегирование и «добычу» данных); 4) анализ данных (позволяющий создавать модели и формулировать теории); 5) визуализация данных (в частности, создание интерфейса для запросов и выдачи результатов обработки данных). Именно такая информационная архитектура использована в проекте CHIA.

Создание глобального архива исторических данных (world-historical data resource) предполагает создание системы хранения, документирования и анализа данных, которая основывается на целостной распределенной модели реляционных баз данных. Для результативности такого проекта необходима постановка четких целей (миссий), которые историки смогут решать с его помощью: 1) компоновка и документирование «больших данных»; 2) создание всеобъемлющего архива; 3) анализ и визуализация данных в глобальном масштабе.

По мнению Мэннинга, создание цифрового архива исторической информации имеет смысл, если данные собираются без ограничений по ме-

сту, времени, теме и объему. Как методологический лозунг такой призыв вполне оправдан, хотя его реализуемость, равно как и качество реализации, — повод для того, чтобы задуматься о кардинальных изменениях в профессии историка¹⁰. Мэннинг признает, что собственно построение глобальной исторической базы данных (базы знаний?) — сложная задача. При этом для успешного решения такой задачи необходимо учитывать опыт работы с историческими данными за последние компьютеризированные полвека. Первое поколение исследователей «больших данных» относится к послевоенному периоду, когда стали формироваться многочисленные, преимущественно статистические, базы данных (к ним можно отнести такие ресурсы, как HRAF, ICPSR, NBER, OECD, WB, а также работы по агрегированию статистических данных Б. Митчелла¹¹). Второе поколение исследователей исторических «больших данных» началось в 1980 гг. и связано с распространением дискуссии о «глобализации» и насущной необходимостью глобальных экономических и межкультурных сравнений (среди проектов этих лет можно выделить UNPOP, IPUMS, GENH, IROWS, а также публикации Э. Мэддисона¹²). Мэннинг полагает, что сегодня настает время для третьего поколения исследователей «больших данных». Третьему поколению суждено проверить глобальные предположения об устройстве мира на разнообразных и по-настоящему всемирных «больших данных» (р. 89–90), причем, помимо типичных исторических индикаторов, уже нет препятствий

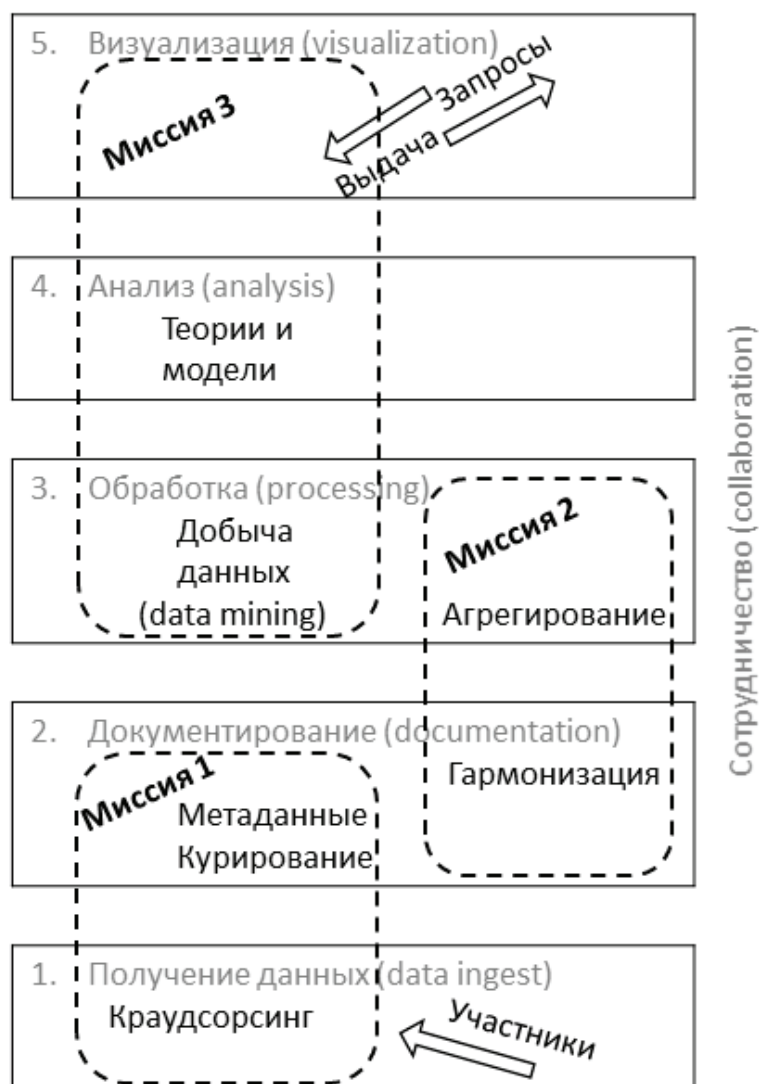
8 CHIA [Electronic Resource]. URL: <http://www.chia.pitt.edu>. В качестве примеров онлайн-проектов по сбору исторической информации также указываются: CLIO-INFRA — clio-infra.eu (Амстердам), CODESRIA — codesria.org (Дакар) и CLACSO — clacso.org.ar (Буэнос-Айрес).

9 Краудсорсингом часто называют привлечение многочисленных волонтеров к трудозатратной деятельности, которую можно выполнить онлайн. См., например: проект по электронному изданию 90-томного собрания сочинений Л. Н. Толстого — «Весь Толстой в один клик». URL: <http://www.readingtolstoy.ru>

¹⁰ Например, Я. Грегори утверждает, что современное поколение историков — первое, которое сталкивается со сложными и неполными цифровыми (или оцифрованными) источниками, кишущими ошибками оцифровки, распознавания, кодирования (см. Gregory I. Challenges and opportunities for digital history. *Front. Digit. Humanit.* 2014. 1:1. DOI: 10.3389/fdigh.2014.00001).

¹¹ Human Relations Area Files [Electronic resource]. URL: <http://hraf.yale.edu>; The Interuniversity Consortium for Political and Social Research [Electronic resource]. URL: <http://icpsr.umich.edu>; The National Bureau of Economic Research [Electronic resource]. URL: <http://nber.org>; Organisation for Economic Co-operation and Development [Electronic resource]. URL: <http://oecd.org>; World Bank [Electronic resource]. URL: <http://worldbank.org>; Mitchell B. R. *International Historical Statistics: 3 Volume Set, 1750–2005*. Palgrave, Macmillan, 2008.

¹² United Nations Population Division — www.un.org/esa/population/; Integrated Public Use Microdata Series [Electronic resource]. URL: <http://ipums.org>; Global Economic History Network [Electronic resource]. URL: <http://lse.ac.uk/economicHistory/Research/GEHN>; Institute for Research on World — Systems [Electronic resource]. URL: <http://irows.ucr.edu>; Maddison Project [Electronic resource]. URL: <http://ggdc.net/maddison>



Основные этапы работы с «большими данными» и связанные с ними миссии ученых-историков Мэннинг

к включению в многомерный статистический анализ климатических или генетических данных.

Вопрос о «больших данных» все больше претендует на важную роль в повестке дня как исторической информатики и «цифровой истории», так и «цифровой гуманитаристики» в целом¹³. При этом весьма уместными кажутся опасения М. Таллера, что «современное представление о цифровых инфраструктурах для гуманитарных наук, кажется, переоценивает идею публикации информации, поэтому инфраструктура для Digital Humanities в ряде

последних дискуссий может быть настолько лишена аналитических соображений, что становится практически неотличимой от цифровой библиотеки (и к тому же не очень сложной)»¹⁴. Именно по этой причине вопрос — как не потерять острый смысл истории в стогу «больших данных»? — становится актуальным методологическим вызовом сегодняшнего дня.

А. Ю. Володин
(МГУ им. М. В. Ломоносова)

¹³ Можяева Г. В., Можяева-Ренья П. Н., Сербин В. А. Цифровая гуманитаристика: организационные формы и инфраструктура исследований // Вестник Томского государственного университета. 2014. № 389. С. 73–81.

¹⁴ Таллер М. Дискуссии вокруг Digital Humanities // Историческая информатика. 2012. № 1. С. 11.