

*К. и. н. И. М. ГАРСКОВА  
(МОСКОВСКИЙ УНИВЕРСИТЕТ)*

***АРХИВИРОВАНИЕ МАШИНОЧИТАЕМЫХ  
ИСТОРИЧЕСКИХ ДАННЫХ:  
международные дискуссии***

Специалисты в области исторической информатики в целом ряде европейских стран уже в течение нескольких лет обсуждают проблемы, свя-

занные с вторичным использованием баз данных, созданных как историками-исследователями, так и «внешними» общественными и государственными организациями. Архивные учреждения (государственные, муниципальные и специализированные) во все большем масштабе начинают сталкиваться с материалами, производимыми этими структурами изначально в машиночитаемом («электронном») виде и представляющими потенциальный интерес для историков. Кроме архивов, возрастающий интерес к проблемам превращения традиционных источников в электронный формат и их публикации проявляют издательские фирмы. Возможности широкого распространения и вторичного использования машиночитаемой информации особенно расширяются с развитием системы компьютерных сетей, подобных internet.

В последние годы в ряде стран Европы делались попытки организации специализированных архивов машиночитаемых исторических данных на базе либо уже существующих архивов МЧД по социальным наукам (Англия, Германия), национальных (Дания), тематических или городских архивов (Швеция), университетов (Норвегия, Голландия, Россия), исследовательских центров (Швеция, Голландия) и т.п.

В рамках программы подготовки и проведения IX Конференции Международной Ассоциации «History and Computing» (IАНС) в конце августа 1994г. в Голландии состоялись российско-голландский семинар по проблемам работы с машиночитаемыми историческими источниками (Амстердам, Институт социальной истории) и круглый стол «Организация архивов машиночитаемых исторических данных в странах Европы».

С голландской стороны в семинаре приняли участие: Л. Брере (Утрехтский университет), Я. ван дер Броек (Гронингский муниципальный архив), Г. Веллинг и Г. Колленер (Гронингский университет) Х. Вальц, Я. Клоостерман и К. Мандемейкерс (Международный институт социальной истории), А.В. Гравер, П. Доорн и Т. Линдبلاد (Лейденский университет), Я. де Минк (Международная организация электронной документации ЮС), Х. Тьялсма и Р. ван Хорик (Голландский Архив машиночитаемых исторических данных), Х. Хофман (Национальный архив Нидерландов), К.Ю. Юргенс (Институт истории Нидерландов). С российской стороны: Л.И. Бородкин, И.М. Гарскова, Н.Б. Селунская (Московский государственный университет), Е.Н. Балькина (Белорусский госуниверситет, Минск), В.Н. Владимиров (Алтайский госуниверситет, Барнаул), С.Г. Кащенко (Санкт-Петербургский госуниверситет), Т.Л. Моисеенко (Институт российской истории РАН) и В.С. Тяжельникова (Московский городской архив). Участие в семинаре принял также и нынешний президент IАНС Я. Олдерволл (Бергенский университет, Норвегия).

Открыли семинар выступления Я. Клоостермана и Х. Вальца, директора и вице-директора Международного института социальной истории,

рассказавших о его работе. Была также организована экскурсия по архиву и отделам Института, познакомившая участников с электронным каталогом, коллекцией фотодокументов и библиотекой.

На первом заседании, которое было посвящено вопросам доступа к машиночитаемым историческим данным, с докладами выступили Л. Бородкин и П. Доорн, рассказавшие о состоянии и перспективах вторичного использования машиночитаемых исторических данных в России и Голландии.

Л. Бородкин, председатель совета Ассоциации «История и компьютер», подчеркнул необходимость значительной финансовой и кадровой поддержки российских архивов, ставящих перед собой задачи хранения и распространения электронных записей в рамках начатой в 1993 г. в РФ Государственной программы информатизации архивов с участием специалистов Росархива, исторического факультета МГУ и Института проблем управления РАН и других организаций. Отметив определенный прогресс в области архивирования машиночитаемых исторических данных в странах СНГ и положительное влияние международных контактов в этой области, Л. Бородкин остановился на основных проблемах работы с электронными записями:

- несбалансированность «спроса и предложения», связанная с пока еще ограниченным использованием информационных технологий в работе историков;
- проблема устаревших форматов данных, большая часть которых существует на магнитных лентах для больших машин со времен внедрения всевозможных АСУ и, несмотря на значительный потенциальный интерес историков к этой информации, практически исключена из вторичного использования;
- несовершенство правовой базы в данной области;
- проблемы доступа к электронным коммуникациям, число пользователей которых среди историков СНГ не превышает двух десятков;
- проблема институционализации Российского архива машиночитаемых исторических данных, существующего пока лишь на общественных началах в виде группы энтузиастов.

В своем докладе П. Доорн, директор Голландского архива машиночитаемых исторических данных (МНОА), характеризуя основных пользователей исторических машиночитаемых данных, подчеркнул, что для историков-исследователей наиболее типичен проблемно-ориентированный подход к созданию МЧД, тогда как для издательских фирм, переводящих в электронный формат и публикующих источники, более важен источник-ориентированный подход (хотя в последнее время им увлечены и многие исследователи). В деятельности традиционных архивов выделяются три основные линии: хранение машиночитаемых данных, поступающих в электронной форме от различных организаций, электронное каталогизирование и непосредственный перевод в электронный формат хранящихся в архивах коллекций - основной целью такой работы является сохранение

источников. «Молодые» архивы машиночитаемых данных в основном занимаются архивированием и распространением коллекций данных, являющихся результатами различных исследовательских проектов, причем их роль важна именно как роль посредников между создателями и вторичными пользователями машиночитаемой информации; в этой роли они способствуют развитию технологии архивирования, документирования и распространения этой информации.

Развитие электронных информационных сетей, соединяющих в себе возможности библиотек, архивов, центров документации и научных центров одновременно, позволяет исследователю получать доступ к информации во всей ее полноте: тестам, цифровому материалу, видео- и аудиоинформации, мультимедиа и др. С точки зрения пользователя такое разрушение институциональных барьеров очень привлекательно, однако задачи, которые при этом решаются разными группами участников этого процесса, хотя во многом и пересекаются, следует четко разделять. Надо определить функции сторон в процессе преобразования данных источников, их хранения и распространения. Сотрудничество с четким "разделением труда" позволит избежать лишней работы и облегчит получение интересных и значительных результатов.

На втором заседании семинара, посвященном обсуждению проблем архивирования и документирования машиночитаемых исторических данных, с докладами выступили И. Гарскова и Х. Тьялсма. И. Гарскова охарактеризовала ситуацию в этой области в России и странах СНГ, начиная с 1960-х гг., когда развитие квантитативной истории подтолкнуло исследователей к созданию и хранению коллекций машиночитаемых данных в форматах, ориентированных на большие ЭВМ. Особенностью этих данных является почти полное отсутствие документации и возможностей вторичного их использования. Период внедрения персональных компьютеров ознаменовал появление значительного числа персональных баз данных - наиболее популярного в настоящее время типа машиночитаемых исторических данных. Персональные БД остаются в своем большинстве проблемно-ориентированными, в связи с чем для их возможного вторичного использования нужно подробно документировать весь процесс преобразования данных из традиционного в электронный формат, учитывая большую роль интерпретации источников исследователем, создающим базу данных.

В этом отношении важной задачей является создание специального центра, который выполнял бы задачи каталогизирования, хранения и распространения МЧД. Пока такой центр (или Банк данных по российской истории) в нашей стране не оформлен юридически, а существует в виде сотрудничества членов национальной АНС, объединенных интересом в этой области. В перспективе деятельность такого центра должна включать информационную поддержку в области архивирования МЧД

(создание электронного каталога описаний баз данных и его публикацию); методологическую и техническую помощь первичным и вторичным пользователям; проверку и оценку качества машиночитаемых файлов; преобразование данных в различные форматы; коллекционирование не только первичных, но и производных или вторичных файлов, равно как и библиографической информации об их использовании; сотрудничество с иными архивными организациями, в том числе международное; изучение возможностей генерализации существующих коллекций МЧД в тематические банки данных.

Говоря о практике архивирования и документирования МЧД в Голландии, точнее, в Голландском Архиве машиночитаемых исторических данных, Х. Тьялсма обратил внимание на сходство в функциях традиционного («бумажного») архива и архива МЧД: главной целью в обоих является хранение данных и их описание. Различия же в области хранения данных, например, заключаются в том, что МЧД надо периодически обновлять, как со стороны физических носителей, так и со стороны программного обеспечения; различия в области документирования связаны, например, с кодировкой исходных данных при их переводе в электронный формат. Все процессы документирования должны быть достаточным образом стандартизированы, причем в этом заинтересованы не только архивы, но и сами разработчики машиночитаемых данных. Стандартная схема описания исторических данных, разработанная в МНОА, признана удачным стандартом описания, так как уделяет большое внимание описанию исходных источников и их связям с производными от них файлами машиночитаемых данных. На основе этой схемы в МНОА создана программа, максимально автоматизирующая для разработчика весь процесс описания данных. Разумеется, в этой области будут возникать новые проблемы, особенно с усложнением форматов МЧД (гипертекст, мультимедиа и т.п.).

Третье, последнее заседание семинара касалось проблем, с которыми сталкиваются уже не исследователи, а профессиональные архивисты, решая задачи автоматизации архивов; на заседании выступили с докладами Т. Моисеенко, В. Тяжелникова и Х. Хофман. Т. Моисеенко в своем докладе подчеркивала, что многие источники в развитых странах доступны как в бумажном, так и в электронном формате, причем первый довольно быстро вытесняется вторым. Однако, пока специфика МЧД как нового вида исторических источников, в частности специфика их хранения, мало обсуждается историками, хотя проблемы для такого обсуждения можно назвать: в чем особенности МЧД как первичных исторических источников, в отличие от традиционных видов; какие характеристики этих данных особенно важны для возможности их вторичного использования; где и как следует хранить МЧД; как хранить и распространять информацию об электронных записях, создаваемых в различных организациях, для обеспечения их повторного использования? Последний вопрос особенно

важен в связи с тем, что до сих пор коллекции данных, хранящиеся в государственных и негосударственных организациях, практически не используются историками, которые либо обращаются к традиционными источниками, либо используют данные коллег, полученные в результате работы над конкретными проектами; о существовании же электронных записей, созданных вне исторического сообщества, исследователи зачастую даже не подозревают, а ведь многие из этих записей находятся под угрозой уничтожения.

Причины, по которым использование «внешних» МЧД историками ограничено или невозможно, Т. Моисеенко делит на две группы: организационные, обусловленные дефицитом места, персонала и денежных средств и архивных организациях и других структурах, связанных с хранением МЧД, неразработанностью законодательства в этой области, недостаточной информацией, - и технические, определяемые жизненным циклом электронных записей, их аппаратной и программной зависимостью и т.п.

Доклад Х. Хофмана был посвящен проекту Национального архива Нидерландов в Гааге по разработке принципов хранения МЧД, поступающих во все возрастающем количестве от различных организаций, в результате чего относительно стабильный и замкнутый мир архивистов превращается в переменчивый мир электронной информации. Создана специальная рабочая группа, в состав которой входят работники Национального архива и муниципальных архивов Амстердама, Роттердама, Гааги и Утрехта, члены Королевского общества архивистов и региональных обществ. Х. Хофман выделил такие достоинства коллекционирования электронных записей, как облегчение процесса оперативной работы с этой информацией (особенно в сфере бизнеса) и отчетности, а также потенциальную обще культурную ценность машиночитаемых данных, как овеച്ചественной памяти общества; именно последнее наиболее интересует архивы. Документы (в том числе и электронные записи) а голландских архивах являются по закону общедоступными, а архивы должны обеспечить порядок их хранения и доступ к ним. Поэтому архивы, помимо хранения и поиска нужной информации, должны выполнять такие функции, как наблюдение за процессом создания электронной информации в государственных организациях, технический, административный и интеллектуальный контроль за этим процессом.

Специальное изучение положения в области архивирования электронных записей в Голландии выявило отсутствие единой государственной политики в области архивирования МЧД, неопределенность статуса таких данных в архивах, так как развитие новых информационных технологий не укладывается в рамки традиционных архивных инструкций; кроме того, децентрализация в хранении электронных записей обнаружила недостаток соответствующего опыта архивирования у лиц, ответственных за хранение этого нового вида информации. Специфика электронных записей

требует не только технического, но организационного и интеллектуального решения указанных проблем, причем последний аспект, т.е. определение того, что и почему надо хранить, определяет и технологию хранения, ее инфраструктуру.

В своем докладе В. Тяжелникова также с точки зрения архивиста оценила положение в нашей стране в области работы с машиночитаемыми данными с 1960-х до начала 1990-х гг. Она указала на интерес, который испытывают архивы к созданию баз данных по материалам, представляющим наибольший интерес для историков-исследователей, и справочной документации. В перспективе видится работа по систематизации машиночитаемых данных, поступающих в архивы как из инициативных исследовательских проектов, так и в результате внедрения информатизации в собственно архивную деятельность; важным представляется создание в структуре архивных органов Центра по вторичному использованию исторических баз данных.

Заседание «круглого стола» (в программе IX Конференции IАНС) по организационным проблемам архивирования машиночитаемых исторических данных, было в основном посвящено обсуждению следующих вопросов: должны ли архивы машиночитаемых исторических данных быть независимыми организациями или входить в другие организации на различных основах?; как наиболее эффективно использовать электронные сети (internet) для сотрудничества между такими архивами?; каковы будущие потребности архивов МЧД в финансировании, персонале и технических средствах?; что будет основным в их деятельности к 2000 году?

В круглом столе под председательством П. Доорна приняли участие; Ш. Андерсен (Британский архив Совета по социально-экономическим исследованиям), Л. Бурнард (Оксфордский центр документации по гуманитарным наукам и искусству), И. Гарскова (Московский университет), Х.-Ю. Маркер (Датский архив машиночитаемых данных), Р. Понеморо (Центральный архив историко-социальных исследований в Германии), Г. Торвальдсен (Норвежский Центр машиночитаемых исторических данных), Х. Тьялсма (Голландский архив машиночитаемых исторических данных), С. Фогельвик (Стокгольмский банк исторических машиночитаемых данных).

Для ответа на поставленные вопросы, подчеркивали выступающие, надо упорядочить некоторые основные понятия;

- что такое «исторический» архив МЧД и когда документ становится историческим источником?;
- что понимается под «данными»: цифровой материал, структурированная алфавитно-цифровая информация, свободный текст, видео-, аудио- и т.п. информация?;

-каково соотношение между строгими правилами моделирования и кодирования социологической информации в соответствующих архивах и стремлением историка сохранить источник в виде, наиболее близком оригиналу, т.е. какова роль источниковедения в деятельности исторического архива МЧД?;

-наконец, как понимать «архив»: как только место для хранения данных или шире - как организацию, оказывающую информационные услуги, и каков круг этих услуг, т.е. должен ли архив машиночитаемых исторических данных ограничиваться хранением и распространением данных, созданных вне его, или активно участвовать в создании как источников-, так и проблемно-ориентированных коллекций МЧД?

Обсуждая эти вопросы, участники подчеркивали, что хотя в идеале архив исторических МЧД может быть независимым, он обязательно будет тесно сотрудничать со всеми заинтересованными архивными, научными и учебными организациями. Основными пользователями такого архива будут участники исследовательских проектов, что накладывает определенные требования к квалификации персонала архива, которая должна включать опыт собственной исследовательской работы. С другой стороны, в круг задач исторического архива МЧД должна входить и такая функция, как предоставление исследователям информации о наличии электронных записей в традиционных (не исторических) архивах и, быть может, самих этих записей, после соответствующих процедур проверки и документирования.

Что касается электронных сетей, в их использовании, наряду с несомненными преимуществами, есть и определенная опасность, состоящая в значительном росте количества плохо проверенных и документированных, ненадежных данных, поступающих в эти сети, в отличие от данных с необходимым уровнем стандартизации и достоверности, которые создаются архивами. Эта стандартизация должна касаться формата, описания, хранения данных и доступа к ним, в связи с чем необходимо продолжение международного сотрудничества национальных архивов для координации своей работы.

В рамках российско-голландского проекта на лето 1995 г. (Москва) намечено очередное заседание международного семинара представителей национальных европейских архивов машиночитаемых исторических данных для продолжения дискуссии по актуальным проблемам их деятельности.