

Конференции**СЕМИНАР****«СОЗДАНИЕ АРХИВА СОЦИОЛОГИЧЕСКИХ ДАННЫХ В РОССИИ:
РЕЗУЛЬТАТЫ И ПЕРСПЕКТИВЫ»****7 декабря 2001 г., Москва****Добрякова Мария Сергеевна****к.с.н., Московское представительство Фонда Форда**Email: mdobryakova@fordfound.org

Не так давно произошло событие, последствия которого могут оказаться весьма серьезными для российского социологического сообщества: в Москве состоялся международный семинар по вопросам создания открытого общедоступного архива социологических данных в России. Особая знаменательность этого события в том, что необходимость создания подобного архива осознавалась и прежде. Попытки продвинуться в этом направлении уже предпринимались российскими исследователями, заинтересованность проявляли также и зарубежные архивы и фонды¹. Однако до сих пор эта идея не была реализована.

Данный семинар был посвящен, прежде всего, презентации проекта Всероссийского центра изучения общественного мнения (ВЦИОМ) по созданию общедоступного архива социологических данных в России (проект осуществляется при поддержке Фонда Форда). Однако затрагивалась и масса других вопросов, связанных с проблемами создания подобных архивов. В частности, российские и зарубежные коллеги рассказывали о своей работе, делились радостями и проблемами на данном поприще. Российские участники представили свои проекты, переданные в архив.

В семинаре приняли участие сотрудники российских и зарубежных университетов, академических институтов, архивов. Среди них В.Артемов (Институт экономики и организации промышленного производства СО РАН), А.Горбачик (Университет Т.Шевченко, Киев), А.Крыштановский (ГУ-ВШЭ, Москва), Ю.Левада (ВЦИОМ), А.Музафаров (РОМИР), Э.Остин [Eric Ostin] (Мичиганский университет), В.Паниото (Киевский международный институт социологии), Н.Растегаева (Институт социологии РАН), А.Суринов (Госкомстат РФ), А.Табуна [Ausma Tabuna] (Латышский архив данных, Рига) Б.Хаусштайн [Brigitte Hausstein] (GESIS, Берлин), К.Шурер [Kevin Schürer] (Британский архив данных, Университет Эссекса), И.Эйновски [Irena Einowski] (Архив социальных наук, Университет Калифорнии, Беркли). И, разумеется, авторы проекта: Л.Косова, А.Пигарев, Л.Хахулина (ВЦИОМ), а также Б.Добров и Т.Юдина (Научно-исследовательский вычислительный центр МГУ).

¹ Можно упомянуть, например, исследовательский семинар по распространению и доступу к научной информации в России и Восточной Европе, организованный 15-18 августа 1997 г. в Университете Эссекса (Великобритания). Среди его участников были представители как западных архивов и родственных им организаций, так и российских. На семинаре обсуждались вопросы распространения информации и увеличения доступа к ней на Западе, сравнивались возможности доступа к данным в различных странах, подробно рассматривалась российская и восточноевропейская инфраструктура данных, технологии распространения данных и др.

После вступительного слова директора ВЦИОМ Ю.Левады **Людмила Хахулина** рассказала об общих подходах к организации общедоступного социологического архива. В частности, докладчица отметила, что идея создания архива витала в воздухе уже с середины 1980-х гг., однако только к концу 1990-х гг. созрели необходимые для этого условия. Во-первых, сейчас существует более десятка организаций, регулярно проводящих опросы общественного мнения. В результате возникают проблемы физической сохранности данных, их стандартизации и обеспечения доступа к ним. Во-вторых, масса исследований проводится на гранты благотворительных организаций, однако у исследователей нет возможности сделать полученные данные доступными более широкой аудитории – а это противоречит принципам благотворительности. В-третьих, расширился круг «потребителей» социологических данных – студентов, аспирантов, преподавателей-социологов, которые сами не имеют возможности производить достаточное количество качественных данных.

Затем **Л.Хахулина, Л.Косова, Т.Юдина** и **Б.Добров** представили проект, выполненный ВЦИОМ в сотрудничестве с НИВЦ МГУ (руководитель Т.Юдина). Оговоримся: поскольку об одном проекте рассказывали несколько людей, близкие вопросы порою затрагивались в разных выступлениях. Поэтому в данном случае представляется более уместным на время отказаться от хронологического изложения в пользу целостного описания проекта, совместив фрагменты разных выступлений.

Итак, проект имеет целью создание российского архива социологических данных, доступных для широкого круга пользователей. Проект был начат год назад, и начальный этап практически завершен. Однако потребуется еще 3-4 года на доработку и развитие определенных направлений, на то, чтобы поставить архив на ноги – с точки зрения и внутренне-рабоче-содержательной, и институционально-организационной.

К настоящему моменту выполнена следующая работа:

1. Собраны массивы первичной информации. Данные отбирались по принципу наибольшей востребованности исследователями и включают общероссийские исследования, сравнительные международные исследования с участием России, электоральные исследования и социологические тренды.

При этом к самим данным также предъявляются определенные требования: они должны быть в формате SPSS, должны содержать описание исследования (имя руководителя, название организации, краткое содержание исследования, ключевые слова), описание выборки, текст вопросника. Заранее оговариваются и условия доступа к информации. Вся информация подготавливается самой организацией, проводившей исследование, при этом ВЦИОМ как создатель архива проверяет эту информацию. Если данные представляют интерес для архива, но не соответствуют указанным требованиям, архив может заплатить владельцу данных за доведение данных до определенного стандарта.

2. Разработана информационно-поисковая система. Сотрудники ВЦИОМ составили рубрикатор, позволяющий классифицировать все данные, поступающие в архив. Группа под руководством Т.Юдиной занимается разработкой системы автоматической индексации текстов и данных.

Доступ к архиву будет осуществляться через Интернет. **Лариса Косова** подробно описала структуру сайта. На нем будет представлена информация о проекте, алфавитный перечень исследований (отдельные исследования, тренды, модули Международной программы социальных обследований [International Social Survey

Program – ISSP) с указанием названия исследования, места его проведения, количества опрошенных респондентов, типа выборки, краткого описания исследования.

Пользователям будет доступен и сам вопросник. Кроме того, если пользователя интересует какой-то конкретный вопрос, можно посмотреть, задавался ли этот вопрос в других исследованиях.

На сайте реализован поиск по ключевым словам (пользователь сможет узнать, в каких исследованиях задавались вопросы на интересующую его тему) и по тексту (поиск осуществляется в тексте вопросов и вариантов ответа в случае закрытых вопросов). Возможно более сложное задание параметров поиска при помощи логических операций «и» и «или».

Разрабатывается также английская версия сайта.

3. Разработан правовой дизайн архива. Подготовлены проекты документов (соглашений) о взаимодействии между владельцем данных, архивом и пользователем. В частности, архиву передается не право собственности на данные, а лишь право распоряжаться ими. При этом право распоряжения может быть полным (когда архив передает данные пользователю, не обращаясь за разрешением к владельцу) либо частичным (когда архив всякий раз запрашивает согласие владельца передать данные пользователю). Данные передаются пользователям на безвозмездной основе в академических целях. Деньги могут взиматься только за материальные носители данных, но не за сами данные.

Предполагаемая процедура работы с архивом такова. Пользователь обращается к сайту архива, находит интересующие его данные, заполняет соответствующую форму и посылает запрос в архив. Архив рассматривает заявку и в случае ее соответствия принципам работы архива высылает пользователю интересующее исследование в формате SPSS.

Авторы проекта подчеркнули, что обеспечение бесплатного доступа к данным имеет в России особое значение. С одной стороны, это вынужденная необходимость, обусловленная весьма ограниченными возможностями российского академического сообщества покупать данные. С другой же стороны, некоторая плата может оказаться, напротив, положительным моментом, стимулирующим потенциальных пользователей четко формулировать свои потребности, а не брать как можно больше данных про запас и на всякий случай.

В числе будущих задач архива – расширение базы массивов первичной информации, отработка технологических и правовых основ работы, информирование российского и зарубежного академического сообщества о своем существовании (тем самым привлечение и владельцев данных, и их пользователей), поиск источников финансирования. Есть также и задача на уровне человеческого (исследовательского) сознания: необходимо сформировать привычку исследователям, с одной стороны, не бояться давать свои данные в архив и, с другой, – обращаться к поиску уже имеющихся вторичных данных прежде, чем приниматься собирать свои. Таким образом, данный проект является по сути поиском модели такого архива².

² Для наших читателей поясним, что доступ к сайту будет открыт приблизительно через четыре месяца (в апреле 2002 г.). Задержка связана с решением организационных вопросов: предполагается, что дальнейшая работа по развитию архива будет передана Независимому институту социальной политики. Архив будет иметь свой собственный сайт, и мы пока не можем указать его веб-адрес, однако заинтересованные наверняка смогут узнать о его

Татьяна Юдина, руководитель группы, занимавшейся в данном проекте разработкой системы автоматической индексации, рассказала также и о своем собственном проекте, тесно связанном с проектом ВЦИОМ. Ее группой, работающей на базе Научного информационно-вычислительного центра МГУ, разрабатывается Университетская информационная система «Россия» – электронная система, интегрирующая данные из различных коллекций. Работа ведется уже почти десять лет, и на настоящий момент базовые направления системы уже разработаны. Система включает коллекции в следующих областях: статистика, карта выборов, бюджетная система РФ и др. Всего система насчитывает более 400 тыс. документов, новые поступления составляют до 500 документов (или 2 Мб) в день. Осуществляется содержательная и техническая обработка ресурсов, их индексация, рубрицирование и аннотирование, возможен сквозной тематический поиск по коллекциям. Подробнее с проектом можно ознакомиться на сайте: <http://www.cir.ru>³.

Борис Добров, также работающий в этой группе, более подробно рассказал о принципах автоматического рубрицирования в системе. Докладчик указал причины, побуждающие осуществлять рубрицирование документов автоматически, – большой объем поступлений, с одной стороны, и субъективизм экспертов, с другой. И затем представил разработанный группой рубрикатор СОФИСТ. Каждая рубрика описывается логическими условиями в терминах тезауруса.

Затем другие участники семинара представили свои коллекции, помещенные в архив ВЦИОМ.

Александр Суринов представлял Госкомстат России. «Для сотрудников Госкомстата», – отметил он, – «научное сообщество – пользователь одновременно самый благодарный и самый неблагодарный». Благодарный – ибо никто не обращается к данным столь активно. И неблагодарный – ибо никого больше не занимают вопросы о том, как эти данные были получены, никто кроме ученых не предъявляет к ним столько требований. Докладчик затронул проблему стандартов. По его мнению, здесь следует быть осторожными, необходимо поддерживать ситуацию, когда мнений несколько. В качестве примера он привел свою область – статистику. В ней существует деление на так называемые жесткие стандарты и мягкие стандарты. Жесткие стандарты используются в отношении данных, имеющих последствия для страны в целом, это некоторый набор обязательных показателей, подлежащих обследованию по единому стандарту. Только в случае соблюдения этих стандартов возможны, например, сравнения между странами и построение определенной политики на основе этих сравнений. Мягкие стандарты с собственным алгоритмом расчета могут использоваться внутри страны, но они не используются для сопоставления с другими странами. А.Суринов коснулся также проблемы платы за доступ к информационным ресурсам. Он считает нормальным, что статистическая информация стоит денег – деньги в данном случае компенсируют стоимость обработки и подготовки информации. Что касается возможностей сотрудничества между Госкомстатом и публичным архивом, то позицию автора можно сформулировать так: «Будем дружить, но с определенными ограничениями». Например, пока не обсуждается вопрос о предоставлении в архив первичных данных переписи населения (как прошлой 1989 г., так и будущей 2002 г.). Причина в том, что данные такого рода имеют большое социально-политическое

открытии на сайте Института: <http://www.socialpolicy.ru>. Прим. ред.

³ См. также нашу рубрику «Профессиональные обзоры», где мы подготовили обзор данной системы, выполненный нашим постоянным обозревателем Е.Новиковой. Прим. ред.

значение (например, они используются для расчетов при определении критериев бедности), и в силу этого доступ к ним правомерен только на федеральном уровне⁴.

Следующий выступавший, **Александр Крыштановский**, отметил важность проекта открытого архива для российского социологического сообщества. Правда, в словах докладчика сквозили некоторые сомнения относительно его жизнеспособности. А.Крыштановский ссылаясь на собственный опыт. Примерно десять лет назад подобный проект был затеян в Институте социологии РАН. Его авторам казалось, что это произведет революцию, все наконец-то станут пользоваться качественными данными. Однако этого не произошло. Может быть, тому виной и наши установки – в качестве примера выступавший привел достаточно распространенную ситуацию при защите диссертации, когда текст диссертации слаб, но в качестве положительного момента отмечается, что «автор провел самостоятельное исследование» (и не важно, что качество его сомнительно). Т.е. проблема существования подобного архива – это прежде всего проблема этики научного сообщества. Другая проблема связана с самими данными: почти никогда не удается решить свою научную задачу с помощью чужих данных. Это возможно, видимо, только на уровне студентов и аспирантов, исследовательские организации будут пользоваться вторичными данными в гораздо меньшей степени. В связи с этим следует более активно внедрять принципы работы с архивными данными в образовательный процесс. Следовательно, имеет смысл несколько иначе взглянуть на саму концепцию архива. Надо изначально понимать, что затраты на его создание не окупятся пользователями – в том числе на уровне обращения к данным, все данные никогда не будут использоваться. При этом количество обращений отнюдь не является критерием качества, здесь другие принципы работы. Тем не менее остается фактом, что архивы – нерентабельны, и необходимо заранее думать о том, как перевести архив на государственное финансирование.

РОМИР представлял **Александр Музафаров**. Переданный в архив проект включает три блока:

- политический мониторинг (в том числе, исследования степени доверия населения Президенту, Правительству, государственным структурам РФ);

⁴ Можно провести любопытное сравнение. Несколько дней спустя, на семинаре по вопросам работы УИС «Россия» Эрик Остин, директор Мичиганского архива, заметил, что наиболее открытым и дружелюбно настроенным государственным агентством по отношению к ним является Бюро переписи населения США [US Census Bureau]. Он отметил, однако, что ситуация не всегда была столь благоприятной. Напротив, в 1960-е гг. Бюро придерживалось политики, аналогичной нынешней политике российского Госкомстата. И только долгие продолжительные переговоры, обращения ученых, давление академического сообщества помогли постепенно изменить ситуацию. В целом, он подчеркнул, что подобные непростые отношения с государственными агентствами существуют во многих странах, и для получения каких-то положительных результатов необходимы действия – и время. В ответ Т.Юдина обратила внимание аудитории на то, что настоящая ситуация в отношении передачи данных Госкомстата в открытый архив – это уже большой шаг вперед по сравнению с той ситуацией, что была несколько лет назад. Так, в 1997 г. на семинаре в Эссекском университете Госкомстат отрицал всякую возможность сотрудничества с открытым архивом, в то время как сейчас, четыре года спустя, в архив уже переданы некоторые коллекции Госкомстата. Видимо, описанная Э.Остином тенденция сближения верна и для России, и остается надеяться, что с течением времени академическое сообщество получит открытый доступ ко всем необходимым данным.

- отношение к событиям социально-политической жизни России (в том числе данные исследований занятости, молодежи, семьи, проведенных в сотрудничестве с Министерством труда и социального развития РФ);
- специальные тематические исследования (например, отношение россиян к зарубежным странам, к странам СНГ, восприятие Интернета и др.).

Исследование проведено в период с октября 2000 г. по сентябрь 2001 г., оно построено на общероссийской репрезентативной выборке, включающей 62 субъекта РФ. Более подробно о проектах РОМИР можно узнать на сайте: <http://www.romir.ru>.

Виктор Артемов представил проект, выполненный Институтом экономики и организации промышленного производства СО РАН. В основе массива – исследование сибирской деревни, начатое Т.И.Заславской. Это лонгитюдное исследование условий жизни, повседневной жизни и бюджетов времени сибирских сел. Выборка квотная для сел и отраслей, и случайная – для респондентов в селе. Всего в выборку включено 16 сел, из них выявлено шесть типов сел в зависимости от степени урбанизации и развитости экономической структуры. Использовался метод ретроспективного опроса о вчерашнем дне. Описано около 300 различных занятий, объединенных в примерно 60 агрегированных видов деятельности. Всего проведено восемь обследований. В архив передан объединенный массив исследований (рассчитаны средние значения по последнему обследованию и разница значений с предыдущими обследованиями, выбраны четыре агрегированные точки – 1975-76 гг., 1986-87 гг., 1993-94 гг. и 1999 г.). Таким образом, это уникальный массив, позволяющий сравнить бюджеты времени селян до перестройки и после.

Нина Растегаева рассказала о проектах Института социологии РАН. Банк данных ИС РАН существует уже достаточно давно, он переживал периоды взлета и падений – застоя. В нем хранятся данные исследований 1968 г. о проблемах молодежи и подростков; результаты исследования под руководством Б.Грушина о Таганроге; исследования проблем девиантного поведения, алкоголизма, наркомании; исследования образа жизни (1982 г., 1986 г.); исследование бюджетов времени В.Патрушева; исследование «Советская женщина сегодня». Нет сомнений, что эти исследования представляют собой большую научную и историческую ценность. Однако передать данные в архив не так просто. В частности, многие данные еще только предстоит перевести в формат SPSS. Кроме того, документация о проектах стала сохраняться только в последнее десятилетие, описания более ранних исследований утеряны. К настоящему времени в архив передано исследование «Современное российское общество в переходный период», проведенное под руководством В.Мансурова. В исследовании рассматриваются вопросы политического устройства России, социального самоопределения респондентов, вопросы коррупции и экологии.

Во второй части семинара зарубежные коллеги рассказали о своем опыте создания архивов. Сами выступавшие шутили, что их презентации удивительно похожи друг на друга – достаточно лишь изменить год и название архива. **Эрик Остин** (Мичиганский университет) рассказал о работе Межуниверситетского Консорциума политических и социальных исследований [ISPSR]⁵ – крупнейшего консорциума архивов социальных данных (архив основан в 1962 г., и сейчас в него входят более 500 университетов и организаций различных стран). **Кевин Шурер** представил Британский архив данных

⁵ Inter-University Consortium for Political and Social Research. Подробнее см. веб-сайт Консорциума: <http://www.icpsr.umich.edu/>.

[UK Data Archive]⁶ (основан в 1967 г.). **Илона Эйновски** рассказала об Архиве данных Университета Калифорнии – который был основан в 1969 г.⁷ Выступавшие достаточно подробно остановились на истории своего архива (опыт, чрезвычайно полезный для совсем молодого российского архива). Обсуждались такие важные вопросы, как стратегии финансирования архивов (в случае указанных западных архивов это (в различных пропорциях): государство/местные власти, членские взносы университетов-партнеров) и стандарты хранения информации и представления данных (так называемый мета-язык DDI, который в настоящее время находится как раз на стадии внедрения).

Бригитте Хаусштайн представляла Центральный архив данных эмпирических исследований в социальных науках (GESIS)⁸, Берлин. Эта организация была основана в 1986 г. и сейчас стремительно развивается. Одно из направлений работы – предоставление доступа к данным, включенным в банк данных Совета европейских архивов социальных данных, международную сеть архивных организаций Европы [Council of European Social Science Data Archives - CESSDA] и Международную федерацию банков данных для социальных наук [International Federation of Data Organizations for the Social Sciences – IFDO]. Б.Хаусштайн рассказала также о ситуации с архивами в странах Восточной Европы и, в частности, о следующих организациях:

- TARKI, Центр информации по социальным исследованиям (Будапешт, Венгрия);
- ESSDA, Эстонские архивы данных социальных наук в Университете Тарту (Тарту, Эстония);
- LZDA, Латвийский архив данных социальных наук (Рига, Латвия);
- ADP, Словенский архив данных социальных наук (Любляна, Словения);
- SDA, Чешский архив данных при Институте социологии Академии наук Чешской республики (Прага, Чехия);
- RSDA, Румынский архив социальных данных, и Институт изучения качества жизни (IQLR);
- Словацкий архив социальных данных при Институте социологии Словацкой академии наук.

Наконец, докладчица представила результаты работы семинара ЮНЕСКО по работе архивов данных социальных наук в Восточной Европе⁹.

В заключение выступили представители Ближнего Зарубежья. После презентаций западных благополучных, сложившихся архивов ситуация с развитием архивов в Украине¹⁰ и Латвии казалась особенно хрупкой, контраст – особенно разительным.

⁶ <http://www.data-archive.ac.uk/>. Подробнее см. обзор Е.Новиковой в прошлом номере журнала (Том 2, № 5, 2001 г.). Прим. ред.

⁷ Подробнее см. веб-сайт архива: <http://ucdata.berkeley.edu>.

⁸ Немецкая инфраструктура социальных наук (Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen; German Social Science Infrastructure Services). Подробнее см. веб-сайт архива: <http://www.social-science-geis.de>.

⁹ Здесь приводятся лишь названия, поскольку мы предполагаем рассказать об этих архивах более подробно в одном из следующих номеров журнала. Прим. ред.

¹⁰ Полезная ремарка в сторону. В.Паниото, представлявший Киевский международный институт социологии, обратил внимание аудитории на то, как правильно говорить: «в

Итак, **В.И.Паниото** и **А.П.Горбачик** рассказали о ситуации в Украине. К настоящему моменту не сохранилось данных ни одного репрезентативного исследования по всей Украине до 1990 г. (В том числе почти не сохранились данные всесоюзной переписи 1989 г.) В 1997-98 гг. началась работа по созданию Банка социологических данных Украины. Получен грант от канадского правительства, работа по созданию банка должна закончиться в течение полугода¹¹. Сейчас на территории Украины работает несколько фирм, собирающих такие данные. Предполагается создать консорциум банков данных. А.Горбачик также более подробно рассказал о формате, в котором хранится бóльшая часть украинских данных (формат OCA), и возможностях его конвертации в формат SPSS.

Аусма Табуна представляла Латышский архив данных. Этот архив только начинает свое существование. Работа по его созданию начиналась буквально с нуля. До сих пор не решены вопросы с собственным сервером, а, следовательно, и с хранением данных. Пакет SPSS по цене пока доступен далеко не всем вузам. Сотрудничают с Гётеборгским университетом (Швеция). Предполагают получить поддержку ЮНЕСКО. В своем выступлении А.Табуна затронула также вопросы установок академического сообщества относительно готовности предоставлять данные своих исследований в открытый архив, а также того, кто же все-таки является целевой аудиторией такого архива.

Таким образом, на семинаре были представлены разные позиции, разные взгляды – можно даже сказать, взгляды из различных временных этапов существования архивов. Как мне кажется, разговор получился не только интересным, но и полезным. И остается надеяться, что российский открытый архив социологических данных будет расти и набирать силы, радуя академическое (и не только?) сообщество.

Украине» или «на Украине». Для современного украинца предлог «на» в данном случае политически некорректен, ибо указывает на политическую несамостоятельность украинского государства. В то время как предлог «в» в данном случае, напротив, подчеркивает независимость и полноценность Украины в качестве государства.

¹¹ См. также сайт архива: <http://qsilver.queensu.ca/~csd/ukarchive/>.

