

**АНДРЕЙ ГОРБАЧИК,**

*кандидат физико-математических наук, доцент факультета социологии и психологии Киевского национального университета имени Тараса Шевченко*

### **Архивы социальных данных: цели существования, формы работы, проблемы создания**

#### *Abstract*

*The author discusses the problems of forming social data archives and the questions connected with the necessity of researchers' special attention to the process of the empirical data documentation. The necessity of developing an integrated software which could support all the stages of empirical social research is emphasized. The experience of social data archive formation and functioning amassed throughout the world is analyzed. The author also discusses the work done at the Institute of Sociology, National Academy of Sciences of Ukraine, in forming a sociological archive and data bank for social studies.*

Закончено эмпирическое исследование, в подготовку и проведение которого вложено много сил и времени. Длительные дискуссии при составлении опросника, тщательная проработка выборки, обучение интервьюеров, контроль и ввод информации. Наконец все данные в компьютере. Можно приступать к самому интересному — анализу собранных данных. Ответы на сформулированные перед началом исследования вопросы, новые вопросы, о которых до начала исследования даже не задумывался, идеи новых исследований... И вот результаты (по крайней мере, первые) опубликованы. Не покидает ощущение, что можно было бы проверить еще пару небольших гипотез, обозначившихся в ходе анализа. Неплохо также сделать сравнение с данными коллег из другого города, об исследовании которых прочел в журнале. Однако нужно начинать новое исследование, и

откладываяешь эти вопросы “на потом”, для аспирантов или для дипломной студенческой работы (а вдруг, в следующем году попадетсЯ хороший дипломник). А что делать с собранными данными? Данными, в которых, по сути, сконцентрировалась значительная часть труда большого коллектива людей, труда интеллектуального и даже физического.

В Англии, например, после написания отчета по проекту приблизительно две трети исследователей хранят собранные в ходе выполнения проекта данные дома или на работе (при этом риск быть разрушенными или утерянными для этих данных очень велик, а нестандартная форма хранения затрудняет доступ к ним для других исследователей), одна пятая исследователей попросту уничтожают эти данные и лишь немногие архивируют их (то есть сдают в какой-либо публично доступный архив). Ситуация в Украине вряд ли чем-то отличается в лучшую сторону.

Подготовка и передача данных эмпирических исследований в архив требуют от исследователя времени и дополнительных усилий. Эти затраты должны быть оправданы. Зачем исследователю передавать собранные и уже проанализированные данные в публично доступный архив? Основываясь на собственном и чужом опыте, можно сформулировать такие основные мотивы:

- публикация данных способствует более широкому обсуждению исследования и, таким образом, увеличивает количество ссылок на исследование;
- вторичный анализ опубликованных данных, произведенный другим исследователем, иным методом, исходящий из отличных теоретических предпосылок, способствует открытой дискуссии по теме исследования, является средством дополнительной проверки выводов и результатов исследования;
- публикация данных способствует тому, что подходы к измерению и методы сбора информации становятся более согласованными (может быть даже, в определенной мере, стандартизированными) в профессиональном сообществе;
- публикация уменьшает дублирование усилий исследователей по сбору одних и тех же данных. Публикация данных крупных общенациональных проектов позволяет вовлекать в анализ этой информации десятки авторов, которые по различным причинам (организационным, финансовым, временным и др.) сами не могут провести подобное масштабное исследование;
- опубликованные данные методически грамотных, профессионально проведенных исследований — важнейший источник информации для студентов и аспирантов.

Однако существует и “обратная сторона медали”. Передавая данные в публичный архив, исследователь должен быть готов:

- к тому, что в его данных и выводах будут искать ошибки. При этом чем интереснее и актуальнее исследование, тем выше вероятность таких поисков;
- к критике со стороны коллег, анализирующих его данные другим методом или с иных позиций;
- к возможному нарушению конфиденциальности собранной информации. Как правило, исследователь обещает респондентам, экспертам и другим “источникам информации” сохранять конфиденци-

альность и использовать полученные данные только в обобщенном “анонимном” виде. А вот человек, подвергающий данные вторичному анализу, таким обещанием непосредственно не связан (хотя, разумеется, существуют общепринятые в профессиональном сообществе этические нормы);

- к возможному нарушению права интеллектуальной собственности на собранные данные.

Тем не менее практика передачи собранных данных в публично доступные архивы в последние годы становится все более распространенной. Многие фонды, выделяя гранты на проведение эмпирических исследований, требуют включать в контракт пункт, в соответствии с которым через определенное время после окончания проекта собранные данные обязательно следует поместить в публично доступный архив. Затраты на подготовку и передачу данных в архив вносятся непосредственно в смету исследования.

Зачастую исследователи не имеют содержательных возражений против помещения данных своих исследований в архив, однако их останавливают трудности подготовки данных к такого рода публикации, особенно необходимость составления подробной сопроводительной документации. Требования, выдвигаемые архивами к такой документации, достаточно высоки, поскольку при отсутствии необходимых описаний данные не могут быть корректно использованы для вторичного анализа. Опыт показывает, что о сопроводительной документации нужно заботиться заранее. Начинать создавать качественную документацию в ходе исследования (а тем более спустя длительное время по его окончании), весьма трудно, а то и просто невозможно.

Чтобы упростить процесс помещения результатов количественных исследований в архив (личный, архив организации, публично доступный архив), уже на самых ранних этапах планирования исследования необходимо сделать следующее:

- составить **полный перечень документов**, которые будут создаваться во время исследования. Он, как правило, включает опросник, описание выборки, инструкции интервьюерам, инструкции кодировщикам, описание файлов данных и другие документы;
- определить **структуру файла данных исследования** (или файлов данных). Необходимо зафиксировать объект (или объекты) анализа. Объектами анализа могут быть “ответы респондента”, “информация о домашнем хозяйстве” и т.п. Отдельные части исследования могут фокусироваться на разных объектах анализа. Следует решить, все ли данные будут представлены в виде одного файла (что можно сделать всегда) или нескольких файлов. Оформление результатов в виде нескольких файлов нередко оказывается более экономным, удобным и логически понятным. Скажем, один файл может содержать результаты интервью всех членов домохозяйств, а другой — информацию о домохозяйствах в целом (жилая площадь дома или квартиры, наличие в домохозяйстве иждивенцев, наличие и размер участка земли и т.п.). Подавая материал из нескольких файлов, необходимо четко определить, каким образом они будут связаны между собой, как найти информацию о каждой конкретной единице анализа. Можно, например, присвоить каждому домохозяйству уникальный номер, который позволит “соединять” результаты интервью отдельных респондентов с информацией об их домохозяйстве;

- установить **общие соглашения о кодировании** данных. Прежде всего речь идет о кодировании отсутствующих значений (“Отказ от ответа”, “Респондент не понимает вопроса”, “Вопрос не задавался” и т.п.) и ошибок (“Ошибка интервьюера”, “Ошибка кодировщика”, “Ошибка респондента” при самостоятельном заполнении анкеты почтового опроса). Необходимо четко определить все возможные такого рода значения для каждой из переменных. Выбор одинаковых (в пределах файлов данных одного исследования) кодов отсутствующих значений и ошибок для всех переменных может представлять трудность вследствие разнородности собираемых данных и специфики используемых шкал измерения;
- установить **соглашения о правилах построения имен** — имен файлов, имен переменных (признаков), имен отдельных объектов. В качестве последних часто выступают уникальные номера. При этом используют как “простую” сплошную нумерацию (вида 1,2,3,...), так и более сложные “структурные” варианты. Примером может послужить следующий способ построения уникальных номеров для домохозяйств: номер домохозяйства всегда записывается в виде семизначного числа, в котором две первые цифры означают номер области, две вторые — номер района в области, три последние — номер домохозяйства в районе. При этом обязательно должен быть определен способ кодирования областей и способ кодирования районов для каждой из областей;
- определить **процедуру проверки данных** во время ввода и интегральной проверки данных после ввода. Спектр проверок может быть достаточно широким: начиная от контроля механических опечаток операторов ввода (например, если переменная принимает только значения 1 или 2, то значение 5 является недопустимым) и заканчивая контролем довольно сложных логических условий “правдоподобности” данных (неправдоподобным, к примеру, выглядит наличие в выборке 17-летнего респондента, имеющего троих детей, старшему из которых 5 лет);
- определить процесс **создания новых переменных** (признаков) и, что очень важно, структуру описания этих новых (вычисленных) переменных. Минимальное описание должно включать формулы вычисления или ссылку на метод и компьютерную программу, с помощью которой производились вычисления. Однако такое минимальное описание не всегда является достаточным;
- определить **структуру полного описания** всех переменных (признаков) файла (или файлов) данных. Описание переменных (“кодировочная книга”, codebook) — один из важнейших, с точки зрения обеспечения возможности вторичного анализа, документов. Опросник никогда не заменит этот документ, поскольку не включает многих кодов значений, возникающих на разных этапах работы с данными, а также кодов открытых и полукоткрытых вопросов. Опросник — это инструмент интервьюера, адресованный респонденту. Описание переменных — это документ, адресованный аналитику, который будет работать с данными.

Для качественных исследований многие задачи выглядят аналогично, хотя, конечно, есть и своя специфика.

Разумеется, перечисленные выше мероприятия не исчерпывают всей подготовительной работы к исследованию. В основном они направлены на

получение логически непротиворечивого (в первую очередь, с точки зрения здравого смысла), целостного и, что очень важно, в достаточной степени документированного набора данных. Именно качественная документация в значительной мере определяет возможность использования данных за пределами коллектива, проводившего исследование (хотя опыт показывает, что качественная документация очень важна и для исследователей, собиравших данные).

Документированность данных существенно зависит от того, какое компьютерное программное обеспечение используется при организации и проведении исследования. Интегрированных программных комплексов, которые бы позволяли согласованно решать типичные задачи эмпирических социальных исследований (как, например, известный комплекс Microsoft Office автоматизирует документооборот и коммуникации в небольшом офисе), пока, к сожалению, нет. Заслуживает внимания набор программных средств Methodologist's Toolchest фирмы Idea Works Inc., представляющий собой девятимодульную экспертную систему, которая помогает составить план исследования и заявку на получение финансирования (включая график работ и бюджет), позволяет определить необходимый объем выборки и метод статистического анализа, дает рекомендации по составлению опросника, методам сбора информации, выбору компьютерного пакета статистического анализа и формы графического представления данных в публикациях и выступлениях. Есть и другие подходы к комплексному охвату процесса эмпирического социологического исследования. Однако эти программные средства пока, к сожалению, не интегрированы. Более того, далеко не все этапы работы исследователя поддержаны в них должным образом.

В своей работе исследователи используют текстовые редакторы, программы для работы с базами данных, пакеты статистического анализа и другие программные средства. В количественных исследованиях центральное место обычно занимает пакет статистического анализа, поскольку возможности именно этой программной системы определяют количества и структуры файлов данных, выбор кодировки признаков, способ контроля вводимой информации, выбор структуры имен и т.п. Новые версии статистических пакетов позволяют интегрировать все больше информации, объединять собственно данные и их описание (документацию), чтобы сделать файл данных действительно самодокументированным. Единого стандарта самодокументированного файла данных в мире пока нет, хотя попытки его создания предпринимаются. Де-факто стандартами (по крайней мере, техническими) в области эмпирических социальных исследований стали обычные текстовые файлы ("обычный текст" или так называемый "формат ASCII/ANSI"), а также файлы известных пакетов статистического анализа данных SPSS и SAS. Именно в этих форматах предпочитают депонировать, хранить и выдавать файлы большинство архивов социальных данных.

В Украине многие социологи, занимающиеся эмпирическими исследованиями, используют в своей работе пакеты статистического анализа данных SPSS (разработка американской фирмы SPSS Inc.) и OCA (разработка автора этой статьи). Файл данных пакета SPSS позволяет хранить достаточно длинные метки переменных (куда можно записать полные формулировки вопросов опросника) и метки значений (куда могут быть записаны предлагаемые респонденту варианты ответов). Весьма гибкой в SPSS является система кодирования и использования отсутствующих значений.

Кроме того, модуль ввода пакета SPSS позволяет контролировать данные во время ввода, облегчает и делает более надежным ввод данных путем программирования переходов и автоматического заполнения значений переменных. В свою очередь, паспорт массива данных в пакете ОСА позволяет хранить длинные названия признаков (переменных), полные формулировки вопросов и предлагаемых альтернатив ответов, текстовые комментарии к массиву в целом (автор исследования, время проведения исследования, описание выборки и т.п.) и к каждому признаку в отдельности. При вычислении нового признака пакет ОСА автоматически документирует процесс вычисления в паспорте. Специальная технология ввода и полуавтоматического кодирования открытых и полуоткрытых вопросов также способствует автоматизации документирования данных. Строгая типизация измерительных шкал для признаков в сочетании с возможностью программирования переходов с автоматическим заполнением значений повышают надежность ввода данных. К тому же, есть возможность организации комплексной интегральной проверки качества введенных данных и организации повторного ввода данных.

Отметим, однако, что, с одной стороны, многие из перечисленных выше возможностей документирования работы с данными во время работы исследователями просто игнорируются, а имеющиеся возможности не используются. Ссылки на отсутствие времени не выдерживают критики и свидетельствуют только о существенных пробелах в профессиональной культуре исследователя. С другой стороны, пока нельзя назвать такую компьютерную программную систему, которая бы решала вопросы сопровождения данных комплексно и согласованно, делая процесс документирования и контроля данных естественным, ненавязчивым, не очень обременительным и в то же время обязательным.

Из-за отсутствия общепризнанных стандартов каждый архив выдвигает свои требования к данным, принимаемым для депонирования. Часто эти требования обусловлены характером деятельности (научной, образовательной) самого архива. Ниже мы сделаем обзор некоторых наиболее известных в мире архивов с точки зрения их деятельности и организации, опираясь в основном на информацию, полученную из сети Интернет. Ориентация на информацию из Интернета определяется не только ее доступностью и оперативностью, но и тем, что “присутствие в Интернете” является самостоятельным косвенным и одновременно очень важным и характерным показателем функционирования архива. Как мы увидим, далеко не все архивы распространяют свои данные, используя сеть Интернет. В большинстве случаев мотивы отказа от такой простой и дешевой формы распространения информации не только и не столько финансовые или связанные с охраной прав на интеллектуальную собственность, сколько “методологические”, продиктованные опасением методологически некорректного использования данных. Присутствие и активность в Интернете прежде всего служит индикатором определенной завершенности формирования архива, подлинной его “публичности”, понимаемой на современном уровне развития технических средств хранения, обработки и передачи данных.

**Центр Роупера** (The Roper Center for Public Opinion Research) — один из известнейших и самых старых архивов в мире. Центр был создан Э.Роупером в 1947 году и находится в университете штата Коннектикут, США. Ныне Центр занимается созданием и сопровождением компьютерного бан-

ка данных исследований общественного мнения, обучением студентов, развитием международных связей в области межнациональных политических и социальных исследований. Здесь хранятся около 13 тысяч наборов данных об исследованиях общественного мнения со всего мира, включая несколько уникальных коллекций (например, все опросы Геллапа, начиная с 1930 года). Большинство данных относятся к США, но есть и чисто “зарубежные” коллекции: данные исследований, проведенных в Японии, в странах Латинской Америки. Центр издает журнал “Public Perspectives”, серию монографий. Существенную помощь в работе Центра оказывают студенты, в большинстве своем работающие как добровольцы.

Списки новых поступлений в архив Центра регулярно публикуются в Интернете. В Интернете свободно (без регистрации и без оплаты) доступен поиск файлов данных исследований в базе данных Центра. Поиск можно производить:

- по фразе на английском языке (поиск вхождения этой фразы в название исследования или в описание исследования);
- по дате исследования (как правило, задается интервал дат);
- по названию организаций, проводивших и финансировавших исследование;
- по характеристике выборки (например, “избиратели”, “молодежь”, “женщины”).

Результатом поиска является уникальный номер файла данных исследования в архиве Центра и общая информация об исследовании. Чтобы получить данные для вторичного анализа, необходимо по телефону или по электронной почте обратиться к сотрудникам Центра.

Центр Роупера имеет архив вопросов, работа с которым осуществляется через систему POLL (Public Opinion Location Library) [1]. В этом архиве единицей хранения является уже не файл исследования, а информация об одном вопросе. Архив хранит более 275 тысяч единиц. Каждая единица хранения включает в себя:

- полную формулировку вопроса и вариантов ответа;
- информацию об исследовании, в котором этот вопрос использовался: дату проведения исследования, кто проводил и кто финансировал исследование, использованный метод сбора информации, описание генеральной совокупности, характеристики и объем выборки и т.п.;
- название тематической рубрики в классификаторе Центра, к которой отнесен вопрос;
- возможно, некоторую дополнительную информацию о вопросе или об исследовании. Это может быть указание на то, что вопрос задавался не всем респондентам, или что таблица отображает объединенные данные нескольких исследований;
- одномерное распределение ответов респондентов на вопрос: либо округленные до целых значений проценты, либо меры центральной тенденции и вариации (среднее, медиана, среднеквадратическое отклонение). Если в исследовании применялось взвешивание выборки, то распределение учитывают вес. Приблизительно для 50 тысяч вопросов из исследований, проведенных после 1990 года, есть двухмерные таблицы, где второй переменной является пол, возраст, образование, доход, регион проживания или раса респондента.

Одноуровневый классификатор POLL включает около 100 тематических рубрик. Отнесение вопроса к определенной рубрике осуществляют сотрудники Центра.

Система POLL позволяет осуществлять поиск единиц хранения:

- по тематической рубрике;
- по дате исследования;
- по названию организации, проводившей исследование;
- по вхождению слова или фразы в текст единицы хранения (так называемый “поиск в тексте по образцу”; есть возможность строить довольно сложные текстовые образцы поиска).

Результатом поиска в POLL является набор одномерных и, возможно, двухмерных таблиц распределений. В ряде случаев этой информации достаточно для работы. Таблицы содержат всю информацию, необходимую для корректной ссылки на исследование. Если нужна более детальная информация или углубленный анализ, то, используя уникальный номер файла данных исследования, который всегда присутствует в таблице, по телефону или по электронной почте можно обратиться к сотрудникам Центра и заказать данные для углубленного вторичного анализа. Можно также заказать проведение анализа сотрудниками Центра. Заметим, однако, что не для всех единиц хранения в POLL имеются исходные данные. В архивах POLL содержится довольно много таблиц, взятых из отчетов и пресс-релизов фирм, проводивших исследование. В таких случаях исходную информацию для вторичного анализа в Центре получить нельзя. Доступ к системе POLL осуществляется только по предварительной подписке через компьютерные сети.

Основанное в 1962 году *Межуниверситетское объединение политических и социальных исследований* (Inter-University Consortium for Political and Social Research, *ICPSR*) содержит и развивает один из крупнейших мировых архивов исследовательских и учебных данных в области социальных наук. Архив содержит данные исследований из разных областей, включая политические науки, социологию, демографию, экономику, историю, образование, геронтологию, криминальное право, здравоохранение, международную политику и международное право. Так, например, архив хранит данные переписи населения в США, начиная с 1790 года. Помимо организации введения, хранения, доступа к компьютерным архивам данных, ICPSR занимается распространением методических материалов, издательской деятельностью, обучением методам количественного анализа данных, оказывает методическую помощь в проведении анализа данных. Членами ICPSR являются более 300 университетов и колледжей Северной Америки, а также сотни образовательных и исследовательских организаций Европы, Австралии, Новой Зеландии, Азии и Латинской Америки. Некоторые тематические коллекции архива ICPSR специально финансируются отдельными фондами. Центр ICPSR находится в Институте социальных исследований университета штата Мичиган, США.

Депонировать свои данные в архивах Объединения могут все желающие, для этого не нужно быть членом ICPSR. Доступ к данным могут получать не только члены Объединения. Общий объем распространенной в 1999 году информации составил более 250 тысяч файлов [2]. Архив ICPSR использует различные способы распространения информации:

- передача файлов данных по сети Интернет. В 1999 году таким образом было разослано около 98% общего объема распространяемой

информации. При этом почти 40% информации востребовано через свободный доступ (бесплатно, получатели, как правило, не являются членами ICPSR). Остальной объем данных получили члены Объединения и те, кто оформил платную подписку;

- рассылка файлов по предварительно оплаченным заказам. Формат данных и носитель информации (дискеты, магнитные ленты, лазерные компакт-диски и т.п.) оговариваются в заказе;
- выпуск и распространение тематических лазерных компакт-дисков и коллекций дискет.

На странице ICPSR в Интернете организован свободный (без регистрации и оплаты) поиск файлов в каталогах архива. Возможен поиск по тексту (вхождению слова или фразы в название исследования или краткое описание исследования), по названию организации, проводившей исследование, по дате исследования. Возможен также поиск по двухуровневому тематическому классификатору. На первом уровне классификатор содержит 18 тематических категорий, на втором уровне — 2–3 категории (обычно это уточнения типа “В США”, “За пределами США”). Описания найденного исследования и его инструмент (как правило, текст опросника) можно свободно получить через Интернет. На странице ICPSR размещены электронные публикации архива и Объединения.

Ежегодно ICPSR проводит в университете штата Мичиган известные во всем мире летние школы по обучению статистике и применению количественных методов анализа данных в социальных науках. Фонды архива часто используются как база эмпирических данных в работе различных курсов этих летних школ.

**Совет европейских архивов социальных данных** (Council of European Social Science Data Archives, **CESSDA**) был создан в 1976 году для распространения данных эмпирических социальных исследований среди европейских исследователей и для образовательных целей. Членами Совета в настоящее время являются архивы из 16 стран Европы. Совет CESSDA активно занимается развитием интеграции европейских архивов данных. На странице CESSDA в Интернете размещается поисковый механизм, позволяющий производить свободный (без регистрации и оплаты) поиск файлов данных в ряде архивов, которые входят в сеть Совета. Поиск можно вести по названию исследования, по названию исследовательской организации, по ключевым словам из реферата исследования, по датам начала и окончания исследования, по названию страны, в которой исследование проводилось. В результате такого поиска выдается информация о том, в каком архиве содержатся данные найденного исследования и общая информация об исследовании.

Одним из крупнейших архивов, входящих в сеть Совета европейских архивов, является **Архив университета в Эссексе**, Великобритания. Архив был создан в 1967 году с целью организации хранения и распространения в электронном виде для вторичного анализа данных в области социальных наук и экономики, полученных из академических, коммерческих и правительственных источников. Архив является крупнейшим в Великобритании хранилищем исследовательских данных в области социальных и гуманитарных наук. В настоящее время он содержит данные более чем 7000 исследований. Для поиска в каталогах архива разработана специальная компьютерная система BIRON. Система BIRON работает в Интернете (поиск не требует регистрации и оплаты) и активно взаимодействует с электронным

тезаурусом терминов в социальных науках HASSET, который, в свою очередь, является средством терминологического контроля при индексировании исследований, хранящихся в архиве. Архивные данные распространяются на различных носителях (включая лазерные компакт-диски, магнитные ленты и дискеты) или через компьютерную сеть. Получение данных из архива возможно только на основе предварительно оплаченного заказа и соглашения. Как правило, цена зависит не от того, какие данные запрошены, а от категории заказчика (отдельный исследователь, образовательное учреждение, государственное учреждение, коммерческая организация и т.д.) и может существенно меняться.

Архив регулярно проводит научные конференции, учебные семинары для пользователей, принимает активное участие в работе летних школ по количественному анализу социальных данных, издает собственный Бюллетень (The Data Archive Bulletin. — Alphaprint (Colchester) Ltd.). Архив в Эссексе тесно сотрудничает с архивом ICPSR.

Одним из крупнейших в Европе считается созданный в 1960 году **Центральный архив в Кельне**, Германия. Этот архив является членом Совета европейских архивов. Центральный архив в Кельне хранит и распространяет для вторичного анализа материалы и результаты эмпирических социальных исследований проведенных в Германии и в Европе. В настоящее время в нем содержатся данные более 8000 исследований. В Интернете организован свободный (не требующий регистрации и оплаты) поиск в каталоге архива. Параметрами для поиска являются название исследования, название организации, проводившей исследование, краткое описание исследования (можно указать слово или фразу), дата исследования, страна, тематическая категория. В соответствии с одноуровневым классификатором выделяется 31 тематическая категория. Сотрудники архива индексируют исследование по одной или нескольким тематическим категориям. В Интернете можно получить полный список исследований, хранящихся в архиве. Для получения хранящихся в архиве данных для вторичного анализа необходимо обращаться к сотрудникам архива.

Из стран, образовавшихся после распада СССР, только Эстония представлена в Совете европейских архивов. **Эстонский архив социальных данных** был создан при факультете социальных наук университета Тарту в марте 1996 года. Архив содержит результаты более 200 исследований, проведенных в Эстонии, начиная с 1975 года. В сети Интернет представлен список этих исследований (в алфавитном порядке и сгруппированный по ключевым словам). Файл данных для вторичного анализа и сопровождающие его документы можно заказать, обратившись к сотрудникам архива.

Исследовательский **Центр информатики социальных исследований TARKI** (Будапешт, Венгрия) был создан в 1985 году при содействии 11 венгерских исследовательских организаций. Центр TARKI является членом Совета европейских архивов. В Центре размещается архив, содержащий данные более 300 исследований, проведенных за последние 15 лет, а также агрегированные данные статистики Венгрии и некоторых стран Европы. Поиск файлов в каталоге архива возможен свободно через Интернет. Для получения файлов данных для вторичного анализа необходимо обращаться к сотрудникам центра.

В Совет европейских архивов также входит **Архив факультета социальных наук** Люблянского университета, Словения.

В 1985 году был организован **Банк социологических данных Института социологии РАН**, Москва. В настоящее время он содержит данные более 500 исследований, проведенных за последние 25 лет в России и в Советском Союзе. В Интернете краткая информация о Банке представлена на странице Института социологии РАН и на сервере немецкой ассоциации GESIS.

В 1998 году фонд “Демократические инициативы” (Киев, Украина) и Центр изучения демократии университета Queen’s (Кингстон, Канада) создали в сети Интернет двуязычный (украинско-английский) **Архив опросов общественного мнения в Украине**. Автор статьи осуществлял техническое руководство украинской частью этого проекта. Архив содержит таблицы одномерных распределений для тринадцати опросов, проведенных Институтом социологии НАН Украины и фондом “Демократические инициативы” в 1993–1997 годах. Все таблицы в архиве проиндексированы с помощью тематического классификатора [3]. Доступ к архиву свободен и осуществляется благодаря стандартным средствам работы в Интернете.

Представленный список архивов ни в коей мере не претендует на полноту. Количество архивов социальных данных различного масштаба исчисляется в мире, без преувеличения, сотнями. Но даже анализ и обобщение изложенной выше информации позволяет сформулировать основные требования к организации и формам деятельности архива социальных данных.

Архив призван *объединять усилия многих организаций*. Для того, чтобы архив функционировал эффективно, его должны поддерживать основные научные и образовательные организации страны, основные производители и потребители социальной информации, информации о результатах эмпирических социальных исследований. Архив должен быть престижным местом депонирования данных. Основная задача его — сохранять, делать доступными для вторичного анализа и распространять данные. Как не может быть полноценной библиотеки без талантливых писателей и без активных читателей, так не может быть архива без тех, кто не жалеет усилий для его пополнения, и тех, кто активно использует материалы архива в своей работе.

Архив должен иметь *стабильные источники финансирования*. Архив не является собственником данных, это всего лишь депозитарий. Архив может взимать плату за предоставление данных, но, как правило, эта плата незначительна и не покрывает текущих расходов. Чаще всего архивы создаются при университетах или крупных исследовательских институтах, оказывающих им финансовую поддержку. Сотрудничество с университетами полезно еще и тем, что позволяет привлекать к технической работе студентов, которые выполняют эту работу в рамках практики или как волонтеры. Финансовыми источниками существования архива могут быть специальные отчисления от бюджетов научно-исследовательских тем (как, например, в Центре Роупера), стабильные финансовые поддержки определенных фондов (фонды могут поддерживать как архив в целом, так и отдельные его тематические коллекции, как в случае с рядом коллекций архива в Эссексе), членские взносы и средства, получаемые за подписку (подобное практикуется в архиве ICSPR), значительная плата за данные, взимаемая с отдельных категорий пользователей (до 1000 фунтов административных расходов с коммерческих организаций в архиве Эссекского университета).

Архив должен постоянно *искать новые источники пополнения своих фондов*, пропагандируя с этой целью свою деятельность и одновременно повышая свой престиж. Источниками данных в основном являются госу-

дарственные и частные исследовательские центры, высшие учебные заведения, органы государственной статистики. Политика пополнения информационных фондов должна быть достаточно гибкой. Архиву следует стремиться принимать разнообразные данные и на различных условиях (от открытого свободного использования и до ограниченного использования исключительно на основе письменного согласия собственников данных). Более того, большинство архивов хранят не только непосредственно первичные данные эмпирических исследований и документацию к ним, но и другую нужную в работе исследователей информацию: данные государственной статистики (данные переписи, агрегированную информацию, необходимую для организации и анализа выборочных исследований), научные публикации, компьютерные программы и документацию к ним, методические материалы.

В дополнение к основной задаче хранения и распространения, для вторичного анализа эмпирических данных архив, как правило, оказывается вовлечен в:

- *издательскую деятельность* — выпуск собственного периодического информационного издания (например, бюллетеня), серий монографий и методических пособий. При этом речь идет как о традиционных изданиях (на бумаге), так и об электронных версиях (в Интернете, на компакт-дисках);
- *научно-образовательную деятельность* — организация курсов и летних школ для молодых ученых, обучение методике работы с архивом, методам количественного и качественного анализа данных, проведение конференций и семинаров;
- *научное обслуживание* — тематический подбор данных, консультирование исследовательских коллективов, помощь в анализе данных. Такая работа обычно выполняется на договорной основе.

Архив должен быть исключительно *надежным местом хранения* данных. Начав работать, он должен постоянно развиваться, модифицироваться и ни при каких обстоятельствах не закрываться. Именно поэтому для архива важен стабильный, пусть даже минимальный, источник финансирования. Кроме того, надежность хранения должна обеспечиваться технически (необходимы надежные современные технические средства, резервное копирование и другие организационные и технические мероприятия).

Архив обязан заботиться об *охране прав на интеллектуальную собственность* авторов исследования, гарантировать *конфиденциальность информации* (прежде всего препятствовать извлечению из массива данных сугубо личной информации о респондентах, за исключением отдельных биографических исследований).

Архив должен заботиться о *качестве хранимых данных*. Требования к принимаемым данным должны быть достаточно высокими. В первую очередь это касается сопроводительной документации, достаточной для проведения вторичного анализа данных заинтересованным специалистом, не являющимся автором исследования. Архив обязан разрабатывать, предъявлять и контролировать требования, и в то же время оказывать всяческое содействие (методическое и техническое) желающим депонировать данные.

Архив должен обеспечивать *гибкий и удобный поиск* (по различным критериям), а также максимально *простой и быстрый доступ* к собираемой информации. И поиск, и доступ следует постоянно совершенствовать, учи-

тывая развитие современных технологий хранения и обработки данных и средств связи. Удобство и легкость доступа не должны противоречить охране авторских прав, обеспечению конфиденциальности информации и надежности ее хранения.

Архив должен поддерживать активные *международные связи*, предполагающие обмен опытом работы и данными с архивами других стран. Именно архив должен концентрировать основные ресурсы (ресурсы данных, методические ресурсы) *сравнительных эмпирических исследований*. И, наконец, эффективная работа архива в современных условиях невозможна без *присутствия в сети Интернет*.

Е. Головаха [3] кратко изложил концепцию и историю создания в Украине социологического архива и банка данных социальных исследований. В настоящее время в Институте социологии НАН Украины ведется отработка концепции организации архива, логической структуры хранения и алгоритмов поиска информации, технических аспектов организации банка данных, разработка методических документов. Создание архива — не «одномоментный» акт. Это достаточно длинный (порой до нескольких лет) процесс становления, а затем — постоянная поддержка и развитие, учитывающие изменение тематики исследований, усовершенствование технических средств сбора, методик измерения, накопления, хранения, поиска и передачи информации и т.п. Поэтому изначальная концепция и соответствующие технические решения должны быть достаточно гибкими и открытыми. Здесь важны мнения различных специалистов, включая тех, для кого создание архива представляет самостоятельную научную и методическую задачу (к ним принадлежит автор этой статьи), и потенциальных пользователей архива, для которых архив должен стать привычным инструментом в повседневной работе.

Предполагается, что архив в первую очередь будет ориентироваться на электронные формы хранения информации, что позволит упростить процесс хранения, сопровождения и распространения данных. К тому же, электронное хранение информации в сочетании с современными компьютерными сетями практически снимает вопрос территориального расположения архива. Предполагается, что архив будет максимально открыт для приема данных. Все методические материалы по подготовке данных к хранению в архиве, необходимые инструментальные (программные) средства будут свободно предоставляться всем желающим депонировать данные.

В архиве будут храниться хорошо задокументированные первичные эмпирические данные, результаты эмпирических исследований, методические материалы. Уже сейчас можно выделить такие составляющие:

- **Банк данных эмпирических исследований.** Единицей хранения в этой части архива будет полностью документированный файл данных с результатами эмпирического исследования, индексированный по тематике исследований. Мониторинговые исследования в банке будут представлены связанной совокупностью файлов. Поиск в этой части архива будет осуществляться по ключевым словам, индексированной тематике исследований, по названию исследования, имени автора исследования, дате сбора данных и другим параметрам, характеризующим исследование в целом. Целью поиска в этой части архи-

ва является получение файла данных и сопроводительной документации для последующего вторичного анализа;

- **Банк вопросов.** Единицей хранения в этой части архива определен отдельный вопрос исследования. Вопросы будут проиндексированы в соответствии с принятым в архиве и постоянно развиваемым многоуровневым тематическим классификатором [3]. Повторяющиеся в нескольких исследованиях вопросы (так называемые “мониторинговые”) будут связаны между собой. Эта часть архива предусматривает поиск по категориям классификатора, текстовым образцам (в полных исходных формулировках вопросов), а также по характеристикам исследования (название, дата проведения, автор исследования и др.). Для выделенных в результате поиска вопросов можно будет получить снабженные всей необходимой для работы описательной информацией таблицы одномерных распределений и, вероятно, двумерные таблицы для некоторых стандартно категоризированных демографических признаков. Создатели архива будут стремиться к тому, чтобы как можно больше файлов исследований были классифицированы до уровня отдельных вопросов, хотя задача эта представляется весьма сложной и трудоемкой;
- **Архив отчетов, публикаций и методических материалов.** В хранящихся в этой части архива полных текстах документов поиск предполагается осуществлять по ключевым словам (индексирующим тематику документа), по текстовым образцам, названию, имени автора, дате и другим характеристикам документа;
- **Статистика.** Прежде всего речь пойдет о данных переписи, текущих данных государственной статистики и другой информации, необходимой для планирования и анализа социологических исследований различной тематики;
- **Терминологический словарь** (профессиональный тезаурус), необходимый как для формирования профессионального языка, так и для индексации тематики исследований и терминологического контроля ключевых слов, характеризующих эту тематику.

Таким образом, архив будет состоять из нескольких постепенно развивающихся частей. Они не будут независимыми, однако каждая из них будет иметь свою систему каталогизации и поиска. Развитие фондов архива может привести к возникновению тематических или региональных коллекций данных. Чрезвычайно интересным и важным представляется развитие архива в направлении хранения данных качественных исследований (кодированных для качественного анализа текстов, аудио- и видеоматериалов), но на начальном этапе эта задача представляется слишком сложной.

Предполагается, что технической платформой для хранения и поиска в архиве данных эмпирических социологических исследований (имеются в виду выборочные, количественные исследования) послужит разработанный автором статьи и уже ставший достаточно популярным в Украине пакет ОСА (версия для работы в среде MS Windows). Получить в архиве файлы данных для самостоятельного вторичного анализа можно будет либо в текстовом формате, либо в формате пакета SPSS, либо в формате пакета ОСА.

Организационные аспекты работы архива (в том числе такие важные, как доступность данных, способы распространения информации, меры по охране авторских прав исследователя и конфиденциальности информа-

ции), наряду с финансовыми аспектами работы архива, требуют широкого обсуждения. Однако уже сейчас можно предположить, что по крайней мере на этапе становления архива информация будет распространяться на основе договора между исследователем, передающим данные, и архивом, а также на основе договора между архивом и пользователем. При этом рассылка информации возможна по электронной почте, на компакт-дисках, дискетах или других технических носителях информации. Подобная практика довольно распространена в мире. Вполне вероятен выпуск тематических компакт-дисков, содержащих как данные, так и необходимую для работы с ними систему поиска.

С самого начала архив будет вести широкую информационную работу, включающую активную деятельность в сети Интернет (электронная страница в сети — наиболее оперативный способ детального информирования о своей работе) и публикацию бюллетеня, пресс-релизов и др. Архив теснейшим образом будет сотрудничать с университетами Украины.

Надеемся, что архив станет если не организатором, то, по крайней мере, неизменным участником научных конференций, школ молодых ученых, “круглых столов” и других мероприятий, собирающих профессиональных исследователей и “потребителей” социологической информации.

Создание архива — это важное событие в жизни всего профессионального сообщества социологов Украины. Архив нельзя рассматривать как проект одной исследовательской группы или даже одного, хотя и весьма авторитетного исследовательского коллектива. Успешность этого проекта будет зависеть от тех, кто сейчас работает над проектированием архива и созданием первого прототипа, и от всего социологического сообщества Украины. Первые должны сделать архив удобным и привлекательным, не допустив дискредитации самой идеи. Вторые — конструктивно отнестись к первым, активно включившись в наполнение и использование такого ресурса. И если для автора этих строк архив представляет самостоятельную научную, а точнее, научную, техническую и организационную задачу, то для большинства людей архив должен стать повседневным инструментом работы, привычным информационным ресурсом, эффективной возможностью (наряду с журналами, сборниками статей и монографиями) ознакомить коллег с результатами собственных исследований.

Автор рассматривает данную статью как приглашение всех специалистов, заинтересованных в создании и развитии социологического архива и банка данных социальных исследований в Украине, к дискуссии на страницах нашего журнала.

### ***Литература***

1. User's Guide to POLL . — The Roper Center for Public Opinion Research. — Storrs (Connecticut), 1999.
2. ICPSR Annual Report 1998–1999. — Ann Arbor (Michigan), 1999.
3. Головаха Е.И. Концептуальные и организационно-методические основы создания “Украинского социологического архива и банка данных социальных исследований” // Социология: теория, методы, маркетинг. — 2000. — № 1. — С. 138–151.