



Направа хрукописи

Стрельникова Анна Владимировна

**ИНФОРМАЦИОННЫЙ РЕСУРС ЭМПИРИЧЕСКОГО ИССЛЕДОВАНИЯ:
ПРОБЛЕМА НЕПОЛНОГО ИСПОЛЬЗОВАНИЯ**

**Специальность 22.00.01 - теория,
методология и история социологии**

**Автореферат
диссертации на соискание ученой степени
кандидата социологических наук**

Москва - 2005

Диссертация выполнена в Государственном университете гуманитарных наук, Институт социологии РАН

Научный руководитель:

доктор социологических наук **Семёнова Виктория Владимировна**

Официальные оппоненты:

доктор социологических наук **Татарова Гульсина Галеевна**

кандидат философских наук **Чернозуб Светлана Петровна**

Ведущая организация:

Социологический Институт РАН (Г. Санкт-Петербург)

Защита состоится « 25 » **МАЯ** 2005 года в _____ часов
на заседании Диссертационного Совета Д.002.011.01
в Институте социологии РАН по адресу: 117218, Москва,
ул. Кржижановского, 24/35, корп. 5, комн. 323.

С диссертацией можно ознакомиться в библиотеке
Института социологии РАН.

Автореферат разослан« 15 » апреля 2005 г.

Ученый секретарь Диссертационного совета,
доктор социологических наук



Семёнова В.В.

ОБЩАЯ ХАРАКТЕРИСТИКА ДИССЕРТАЦИИ

Актуальность исследования

В отечественной социологии сложился стереотип, что «проведение исследования» - это, прежде всего, выход в поле и самостоятельный сбор первичных данных.¹ В результате мы наблюдаем парадоксальную ситуацию: в России более полувека проводятся многочисленные социологические исследования, и уже должны быть накоплены огромные массивы эмпирического материала, отражающего особенности жизни социума за многие годы. Но при этом информационные ресурсы, которые можно использовать для вторичного анализа, практически не представлены в практике социологических исследований, отсутствуют также и доступные хранилища. Это приводит к тому, что социология остается преимущественно «полевым» феноменом, а ресурс исследовательских данных вырабатывается не полностью.

Обращение к проблематике неполного использования социологических данных особенно актуально сейчас, в условиях перехода к рыночным отношениям и коммерциализации науки. Исследовательский информационный ресурс стал одной из разновидностей товара на высоко-конкурентном и насыщенном рынке, а использование экономических понятий становится востребованным не только в социологии, но и в других областях гуманитарного знания. Там, где возникают вопросы конкуренции, затрат и качества, должен возникнуть и стимул решать научные задачи эффективнее, используя накопленные вторичные данные.

Изучение проблемы неполного использования социологических данных в контексте соотношения «первичных» и «вторичных» исследований приводит к анализу не только методических, но и методологических и институциональных причин такого дисбаланса, а также к рассмотрению закономерностей развития социологии как научной дисциплины. Несмотря на то, что срезовый выход «в поле» дает ответы на конкретные исследовательские вопросы, в условиях избытка эмпирики обостряется научоведческая проблема: в какой мере указанный тип исследований обеспечивает полноту научных выводов; может ли он служить целям объяснения и понимания долгосрочных тенденций в обществе; можно ли без анализа вто-

¹ Воронков В. За пределами публичного пространства (рефлексия социолога)// Неформальная экономика. Россия и мир /Подред. Т. Шанина М.: Логос, 1999. С. 85.

ричных данных понять, какие из противоречащих научных результатов верны, а какие ложны. Так как науку делает наукой способность объяснять, сравнивать, сопоставлять, то необходимо признать, что этап преимущественно эмпирического сбора данных является «подготовительным» этапом для формирования научной дисциплины или направления.²

Слабое использование вторичных данных порождает целый ряд проблем и препятствий в развитии социологической науки и научного сообщества. Среди них - неразработанность многих перспективных методов сбора и анализа социологической информации,³ снижение уровня культуры использования математико-статистических методов в социологии,⁴ разобщенность исследователей и утрата преемственности знания, а также ряд других проблем.

Переизбыток эмпирических исследований (в противовес аналитическим, обобщающим, сравнительным, методическим) начинает осознаваться как «затоваривание». И основной способ преодоления этого «затоваривания» - обращение к вторичным данным.

Указанные причины обусловили актуальность диссертационного исследования.

Степень разработанности проблемы

Социология как научная дисциплина может быть представлена на нескольких уровнях познания. Традиционно они выделяются по оси «фундаментальное-прикладное»: уровень парадигм, уровень больших теорий, уровень частных концепций, предметный уровень. Но каждый из этих уровней также является дискретным. На предметном, или эмпириическом, уровне можно выделить следующие ключевые этапы: определение *методологической основы исследования, поиск, сбор, анализ, хранение* данных.

По теме диссертационной работы накоплена определенная литература, касающаяся общих и частных вопросов *методологии социологических ис-*

²Collins K Sociology: prescience or antiscience? // American Sociological Review. 1989. Vol. 54. P.124-139.

* Маслова О.М., Толстова Ю.Н. Методология и методы//Социология в России/Под ред. В.А. Ядова. М.: Издательство Института социологии РАН, 1998. С. 70-104.

⁴ Как правило, для использования вторичного анализа требуется серьезная подготовка в области статистических процедур и специальных программ; практически невозможно ограничиться простейшими описательными статистиками - они уже подсчитаны тем исследователем, для кого эти данные были первичным материалом.

следований, а также характеристик, особенностей и предназначения социологической информации. Хорошо развит также сегмент работ по анализу социологических данных. Однако вопросы использования вторичных данных, поиска и хранения социологической информации, освещены в литературе в значительно меньшей степени. Остаются практически нераскрытыми и критерии эффективного функционирования научного сообщества, как нормативной модели современной социологии знания.

Несмотря на то, что впервые вопросы об эффективном использовании исследовательских данных и результативной организации научной деятельности были поставлены еще в работах Р. Мертена и его учеников,⁵ исследований на данную тему в современной литературе немного. Анализ источников свидетельствует, что напрямую указанная проблема практически не затрагивается, или упоминается только в контексте возможностей для интерпретации (И.А. Бутенко, В.Н. Карпович).

Среди работ, посвященных теоретико-методологическим проблемам социологии, общим и частным вопросам методологии социологических исследований, необходимо выделить труды Г.С. Батыгина, М.Блауга, П.Бурдье, М.Буравого, И.А. Бутенко, Э.Гидденса, А.Б.Гофмана, А.А.Давыдова, И.Ф. Девятко, В.В. Ильина, С. Квале, П. Лазарсфельда, ИЛакатоса, О.М.Масловой, КЛоппера, Дж. Тернера, П.Уинча, В.А. Ядова. Указанные авторы детально описывают сильные и слабые стороны различных теоретико-методологических направлений; анализируют причины и следствия стратегического разделения на качественный и количественный подходы, прорабатывают различные стратегии и методы исследования, ведут дискуссию о роли исследователя и респондента, о формах проведения и организации исследований, описывают методики повышения надежности и различные критерии проверки данных, поднимают проблематику интерпретации социологических данных. В нашем исследовании этот сегмент работ необходим для анализа возможных методологических препятствий к полному использованию данных.

Среди авторов, работающих в направлении анализа социологических данных, следует выделить Х. Критцер, А.О. Крыштановского, К. Марш,

⁵Merton K The Matthew Effect in Science // Science. Vol. 159 (3810), 1968. P. 56-63; Merton R. The Sociology of Science: An Episodic Memoir // The Sociology of Science in Europe / Ed. by R. Merton and J. Gaston. Southern Illinois Univ. Press, 1977. P. 3-144; Merton R The Sociology of

Г.Г.Татарову, Ю.Н. Толстову, Н.С.Ростовцева, и др. Проработкой стратегий анализа качественных данных занимаются К. Гирц, В.Б. Голофаст, А.С.Готлиб, Т. ван Дейк, Н. Дензин, Е.Ю. Мещеркина, К.Пламмер, В.В.Семенова, Э. Страусе, Х. Уолкотт, А. Эббот, и др.

Однако почти все перечисленные авторы, заводя речь об анализе данных, описывают ситуацию «после поля». Они ориентированы на исследователя, который провел полевые работы, собрал данные, и теперь не знает, что с ними делать, как анализировать и интерпретировать. Несмотря на широкий тематический репертуар (использование как серьезных математических процедур, так и описательных приемов; систематизация способов организации научного текста; дискуссии о пределах для интерпретаций в социальной теории), вторичный анализ, как стратегия привлечения уже имеющихся до «поля» данных, удостаивается лишь кратких упоминаний. Исключение составляют работы таких авторов, как С. Арбер, Э.М. Коржева, Л. Корти, К.Кэрвайт, Г.И. Саганенко, П.Томпсон, О.Е.Трущенко, Дж. Фаустер. Указанные исследователи проблематизируют использование вторичных данных и задаются вопросом расширения сферы поиска информационных источников, применения архивных ресурсов. Особо следует отметить английского исследователя П.Томпсона и его коллег, которые, выпустив в 1998 г. полемическую статью «Почему вы сидите на своих качественных данных?»,⁶ положили начало дискуссии о возможности многократного использования данных, полученных «качественными» методами. Однако российские ученые до сих пор практически не вовлечены в эту дискуссию.

Особую группу западных авторов составляют исследователи, чьи публикации и научные интересы изначально посвящены работе с вторичными материалами. Это, в первую очередь, М.Хэним, А.Дэйл, Дж.Дункан, М.Проктор, Дж.Эддер.⁷ Отдельным направлением является также Data Mining (букв, «извлечение данных») - методология использования больших

⁶ Corti L.Thompson P. Are you sitting on your qualitative data? Qualidata's mission // International Journal of Social Research Methodology, 1998.Vol. 1 (1). P. 85-89.

⁷ Dak A., Arber S., Proctor M. Doing secondary analysis//Contemporary social research series. L: Unwin Hyman Ltd, 1980. № 17; Methods of life course research: Qualitative and quantitative approaches/ Ed by J.ZGiele, G.H.Elder.L:Sage Publications, 1998; Duncan G.J., Kalton G. Issues of design and analysis of surveys across time // International statistical review. 1987. Vol. 55. P. 97-117. Hakim C Secondary analysis in social research. A guide to data sources and methods with examples. L: George Allen & Unwin, 1982.

массивов вторичных данных. Проблематика data mining, от круга приоритетных исследовательских задач до особенностей практического применения, нашла детальное отражение в работах таких зарубежных авторов, как Р. Блэкхаус, Т. Майер, С. Перес, Дж. Перл, К. Хувер.⁸

Теоретико-методологические основы диссертации

Диссертация основана на концептуальной проработке общесоциологической методологии в контексте использования исследовательских данных (Г.С.Батыгин, И.А.Бутенко, К.Кэрвайт, П.Лазарсфельд, П.Томпсон, Г.И. Саганенко, М. Хэким), и «частных» исследовательских методологий, таких, как методология анализа данных (Г.Г. Татарова⁹). В диссертационном исследовании под *информационным ресурсом* подразумеваются все данные (материалы), на основе которых в социологии можно решать соответствующие теоретико-методологические или прикладные задачи, и уточняется, что информационный ресурс может быть представлен первичными и вторичными источниками данных. Работа с вторичными данными подразумевает, что используется уже «готовая» информация, которая была собрана и обработана кем-либо из предшественников.

Ключевое понятие, используемое в работе - *неполное использование* исследовательского информационного ресурса, под которым автор понимает незаслуженное исключение из анализа тех данных, которые могли бы предоставить необходимую информацию. Сюда относится и ситуация поверхностного первичного анализа, и ситуация невостребованности данных со стороны других социологов. В диссертационном исследовании проводится разделение на теоретический (схемы анализа, гипотезы и концепции) и эмпирический (массивы числовых данных, тексты интервью) сегменты информационного ресурса, и далее рассматривается неполное использование *эмпирического* информационного ресурса.

Постановка проблемы эффективного использования информационных

⁸ Backhouse R., Morgan M. Introduction: is data mining a methodological problem? // Journal of Economic Methodology. 2000. N 7. Vol. 2, P. 171-181; Hoover K., Pearl J. Causality: models, reasoning, and inference. Cambridge: Cambridge university press, 2000; Perez S. Three attitudes towards data mining // Journal of Economic Methodology. 2000. N 7. Vol. 2, P. 195-210; Mayer T. Data mining: a reconsideration // Journal of Economic Methodology. 2000. N 7. Vol. 2, P. 183-194.

⁹ См., например: Татарова Г.Г. От постулатов эмпирической социологии к методологии анализа данных // Социология: 4 М., 1999. № 11. С. 51-71.

ресурсов и эффективной организации научной работы заставляет обращаться к практике научоведческих исследований, к социологии науки, опираясь на работы зарубежных авторов: П.Бергера и ТЛукмана, Д.Бока, Дж.Гилберта, М.Малкея, Р.Мертона, Р.Коллинза, Т.Куна, Т.Парсонса, Д.Прайса; и российских исследователей: П.Г.Арефьева, РЛ.Винклер, ЮЛ.Качанова, В.Ж.Келле, Э.М.Мирского, Е.З.Мирской, В.В.Налимова, В.С.Степина, Р.Фрумкиной, А.В.Юревича и др. Работы указанных исследователей позволили автору рассмотреть нормативную модель функционирования социологии как научной дисциплины и применить ее к проблематике неполного использования социологических данных, на теоретическом уровне - посредством изучения связи между особенностями использования информационного ресурса и различными индикаторами, характеризующими современное состояние социальной науки, и на эмпирическом уровне - посредством разработки и подсчета индекса неполного использования.

Объектом диссертационного исследования являются *вторичные данные как информационный ресурс в социологических исследованиях*. В качестве **предмета** исследования выбраны особенности использования информационного ресурса: *полное и неполное использование*.

Целью диссертационного исследования является комплексный анализ проблемы неполного использования информационного ресурса в социологических исследованиях, и выработка эффективных моделей многократного использования данных.

Задачи исследования

Для достижения цели исследования в диссертационной работе предусмотрено решение следующих задач:

1. Рассмотреть проблему вторичного информационного ресурса в социологических исследованиях на трех теоретико-методологических уровнях: научоведческом, общеметодологическом и на уровне частной методологии (анализ данных).
2. Выявить предпосылки неполного использования социологических данных на каждом из уровней и построить «дерево проблемы».
3. Систематизировать понятия и разработать концепцию полного/неполного использования исследовательских данных.
4. Охарактеризовать влияние структурных, институциональных факторов функционирования научного сообщества на исследовательские при-

емы работы с данными.

. 5. Предложить способы решения проблемы неполного использования исследовательских данных.

Эмпирическая база исследования формировалась из вторичных источников информации. В анализ были включены сведения о 30 исследовательских архивах,¹⁰ в том числе российские: «Банк данных социологических исследований», Единый архив социологических данных «Софист», архив ЦНГИ (Центра независимых гуманитарных исследований), INSYS, ЭБД (Электронная библиотека диссертаций при Российской государственной библиотеке), архивы исследовательских организаций ФОМ, ВЦИОМ, ряд ресурсных центров; и зарубежные: The Roper Center For Public Opinion Research, German National Survey Research Centre, Terman Study Center at Stanford, INFRATEST, Berkeley Study Archive, Oakland Study Archive, Research Center at Radcliffe College, North Carolina State Data Center, и др. Для исследовательского анализа также были привлечены статистические данные Госкомстата РФ. Измерение индекса неполного использования осуществлялось посредством анализа публикационного массива журналов «Социологические исследования» (180 статей¹¹), «International Journal of Social Research Methodology» (20 статей), «Oral history» (36 статей). Также в диссертационной работе привлекались учебные программы по анализу социологических данных.

Научная новизна диссертационного исследования заключается в том, что в рамках общесоциологической методологии и в традиции научно-важеского подхода получили дальнейшее развитие и конкретизацию исследования важной научно-практической проблемы - неполного использования социологической информации.

Научная новизна выносимых на защиту теоретических и методических положений состоит в следующем:

1. Комплексный анализ проблемы использования информационного ресурса впервые проведен с позиций научноведческого подхода. Проанализирована проблема хранения и воспроизведения данных как функция научного знания.

¹⁰ Как традиционных, так и электронных.

¹¹ За период 2000-2004 г.г.

2. Доказано, что проблема неполного использования исследовательских данных связана с такими научоведческими индикаторами, как однородность-фрагментированность социологических теорий и концепций, разобщенность-преемственность в передаче знания, соотношение стадий накопления-систематизации, норма-аномалия в исследовательской культуре, независимость-коммерциализация исследований.

3. Классифицированы понятия, принципы и методы использования информационного ресурса. Уточнен и дополнен концептуальный аппарат, связанный с использованием вторичных социологических данных.

4. Разработан и апробирован на массиве социологических публикаций индекс неполного использования социологических данных.

5. Сформулированы методологические аспекты проектирования научного сервиса, связанного с хранением и доступом к вторичным исследовательским данным. Рассмотрена роль исследовательских архивов в развитии социологических направлений, систематизированы сведения о наиболее известных и успешно действующих исследовательских архивах.

Основные положения, выносимые на защиту

1. У всех исследовательских данных, если они получены с соблюдением соответствующих методологических правил, имеется ресурс для неоднократного использования. Однако на современном этапе российской социологии этот ресурс используется недостаточно, что порождает группу проблем и препятствий как для развития отдельных исследовательских направлений, так и для нормального функционирования научного сообщества в целом.¹²

2. Проблема неполного использования социологических данных связана со следующими *методологическими* факторами: рассогласованность надметодологических и методологических установок, парадигмальные ограничения (крайние формы эмпиризма и крайние формы гуманистического подхода), дисбаланс в предварительной («доэмпирической») и завершающей («пост-эмпирической») стадиях исследования, и со следующими *институциональными* факторами: фрагментированность социологического знания; несоблюдение принципа преемственности; дисбаланс в стадиях накопления и систематизации данных; пробелы в исследовательской культуре, коммерциализация социологических исследований.

¹² В частности, экспортный потенциал российской социологии имеет преимущественно «сырьевой» характер: в международных проектах российская сторона чаще всего выступает в качестве поставщика первичного научного «сырья», а подготовка самого научного «продукта» происходит за рубежом.

3. Индекс неполного использования социологических данных в российских исследованиях является высоким как в абсолютном выражении, так и в сопоставлении с западными работами. Модальное значение индекса, по итогам анализа публикационного массива, равно 3 (при размерности индекса от -5 до 5). При этом собранный «в поле» материал оказывается большим по объему, нежели первоначально поставленные цели, и поэтому может использоваться другими исследователями.

4. Уровень обеспеченности российских социологов исследовательскими архивными учреждениями и степень вовлеченности в практику их использования являются низкими. Национальный архив только начинает формироваться, локальные архивы поддерживаются незначительным числом научно-образовательных учреждений, а те, что есть - не объединены в единую сеть. Архивы негосударственных исследовательских организаций предоставляют только ограниченный доступ к данным, обычно на коммерческой основе или по личным контактам, что делает их малодоступными.

5. Дальнейшее развитие социологии как области научного знания, основанной на сравнительном анализе, возможно только при повышении статуса вторичных исследований в социальных науках и налаживании системы хранения и доступа к массивам проведенных исследований, а также при повышении уровня профессиональной подготовки в области предоставления и поиска необходимых информационных ресурсов, в области методологии анализа данных и в частности, в направлении Data Mining.

6. Исследовательские архивы должны формироваться по единым стандартам, с последующим объединением в межвузовские предметные или отраслевые базы данных. Передача в общедоступный банк данных результатов исследований должна, с одной стороны, стать массовой, и с другой - сопровождаться сохранением авторских прав и оговариваться условиями о различных уровнях допуска к использованию информационного продукта.

7. Сетевой подход к преодолению неполного использования вторичных данных связан с развитием межнаучной коммуникации, с оформлением «незримого колледжа»,¹³ который будет разрабатывать указанное направление.

¹³ Представление о научных коммуникациях как о «незримых колледжах» было введено в научковедческий оборот Д. Берналом и развернуто Д. Прайсом. Позднее оно было подвергнуто неоднократным эмпирическим исследованиям (Д. Крэйн, Н. Маллинз, Б. Гриффит и др.), в которых была доказана его опериональная значимость.

Практическая значимость результатов диссертационного исследования

Практическая ценность исследования заключается в предложении мер для повышения эффективности работы с исследовательскими данными. Результаты диссертационного исследования могут послужить основой для повышения статуса вторичных данных и расширения сегмента информационного сервиса в социальных науках, связанного с хранением, систематизацией и поиском данных.

Результаты работы также могут применяться в качестве аналитической информации о перспективах и преимуществах работы с различными западными и российскими исследовательскими архивами. Материалы диссертационного исследования могут быть также использованы в преподавании учебных курсов по методологии социологических исследований, по вторичному анализу социологических данных.

Апробация работы

Основные результаты диссертационного исследования были представлены на Всероссийской научной конференции «Научный сервис в сети Интернет» (20-25 сентября 2004 г., г.Новороссийск), на круглом столе фонда Фридриха Эберта (19 ноября 2001 г., г. Москва), на конференции молодых ученых в Санкт-Петербургском Государственном Университете (27-28 октября 2004 г., г. Санкт-Петербург), на совместном заседании группы «Социодинамика поколений», сектора «Банк данных социологических исследований» и Центра исследований социальных трансформаций Института социологии РАН (10 декабря 2004 г., г. Москва), в специализированных научных Интернет-сообществах, опубликованы в научной печати. Положения диссертационной работы обсуждались на первой сессии международной научно-практической программы «HESP regional seminar for excellence in teaching» ее экспертами и участниками (9-29 августа 2004 г., г. Львов).

Структура диссертации

Диссертация состоит из введения, четырех глав, заключения, библиографии и приложения.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во *Введении* обосновывается актуальность темы, показана степень разработанности проблемы, определены объект, предмет, цели и задачи диссертационного исследования.

В первой главе, «*Неполное использование исследовательских данных как социологическая проблема*», рассматриваются основные теоретико-методологические подходы к изучению проблемы неполного использования исследовательских данных, вводится и уточняется понятие «информационный ресурс», анализируются сущность и предпосылки проблемы неполного использования исследовательского информационного ресурса с точки зрения формирования социологии как области научного знания.

В параграфе 1.1., «*Теоретико-методологические подходы к изучению проблемы использования исследовательских данных*» определяется, что изучение особенностей использования исследовательских данных возможно на трех теоретико-методологических уровнях: общеметодологическом, методологии анализа данных и научоведческом; сформулировано обобщающее понятие «информационный ресурс эмпирического исследования».

В диссертационном исследовании информационный ресурс понимается как теоретическая и эмпирическая база для обоснованных научных выводов, и рассматриваются возможности для вторичного использования эмпирического информационного ресурса, то есть тех «сырых» данных, за которыми один исследователь направляется в поле, и которые затем могут использоваться для решения новых задач.

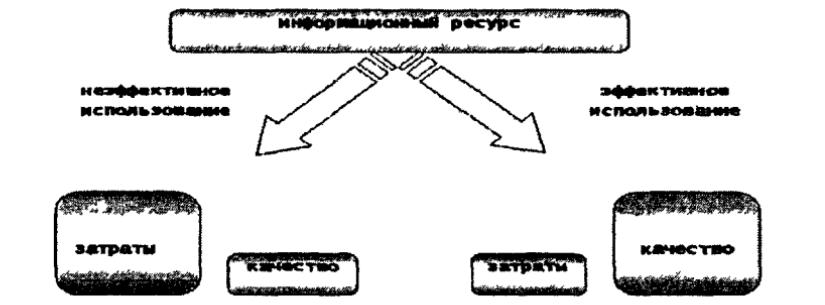
К институционализации дисциплины приводит установление системы дифференциации, интеграции, и воспроизведения знания. С одной стороны, воспроизведение знания подразумевает, что накопленный информационный ресурс не извлекается из научного оборота, а используется многократно. С другой стороны, работы предшественников нуждаются в изучении с новых перспектив, в свете новых научных фактов. Следовательно, для развития научной дисциплины необходимо поддерживать определенное соотношение первичного и вторичного информационного ресурса. В случае дисбаланса этих ресурсов возникает проблема неполного использования социологических данных.

Неполное использование исследовательского информационного

ресурса означает незаслуженное исключение из анализа тех данных, которые могли бы предоставить необходимую информацию. Это и ситуация поверхностного первичного анализа, и ситуация невостребованности данных со стороны других социологов (однократное, «срезовое» использование). Полное использование исследовательского информационного ресурса означает многократное, повторное обращение к данным, работу одновременно и с первичными, и с вторичными источниками. Развивая идеи И. Лакатоса,¹⁴ автор утверждает, что социологичность должна подразумевать интегрированность выводов в эмпирическую историю проблемы. Возможность анализа вторичных данных по отношению к актуальной научной проблеме является базой для обоснованной социологической теории и социологических прогнозов, для выявления случайных или концептуальных противоречий в результатах различных исследований.

Говоря об исследовательских данных как об информационном ресурсе, очерчивается круг для дальнейших рассуждений в терминах затрат, эффективности, качества (см. Рисунок 1).

Рисунок 1. Использование информационного ресурса



Представление о научном результате как товаре сложилось не сразу и до сих пор не является общепризнанным. Однако именно конкурентоспособность научных результатов стала критерием пересмотра условий государственной монополии на научное производство, а эффективность постепенно стала признаваться значимым индикатором в использовании информа-

¹⁴ Лакатос И. Фальсификация и методология научно-исследовательских программ. М.: Медиум, 1995.

мационного ресурса, причем как для тех, кто производит этот ресурс, так и для тех, то потребляет.

В диссертации определяется, что у всех данных (и в особенности - у текстовых) имеется вторичный ресурс. Этой же точки зрения придерживаются и американские исследователи Дж.Лоуб, Р.Сэмпсон и С.МакКей.¹⁵ Российские авторы (напр., Ю.Н. Толстова) также рассматривают один из вариантов исследовательской стратегии, в которой ученый отказывается от идеи *конечности* исследования. Это подразумевает не только готовность к постоянному возврату к одним и тем же данным, но и активное вовлечение новых источников с целью уточнения понимания происходящих процессов.¹⁶ Указанная исследовательская стратегия может быть описана с помощью метафоры герменевтического круга. Особенностью неоднократного использования данных является прохождение полного цикла (круга) интерпретаций, после которого может начаться новый виток интерпретаций, с другой исследовательской перспективы.

В параграфе 1.2, «*Институциональные ограничения для полного использования информационного ресурса*», определяются основные предпосылки для неполного использования данных с точки зрения формирования социологии как области научного знания. Поставленная проблема анализируется посредством следующих научоведческих индикаторов: соответствие императивам, которые регулируют устойчивое функционирование научного сообщества, парадигмальная согласованность (теоретико-методологическая однородность), стадия развития научного направления, наличие аномалий в воспроизведстве социологического знания и в функционировании научного сообщества.

Р. Мертон одним из первых предложил теоретическую схему науки как социального института, регулятором которого выступает *научный этос* – совокупность особых норм. Возникла нормативная модель функционирования научного сообщества, включающая императивы универсализма, коллективизма, бескорыстности и организованного скептицизма.¹⁷

¹⁵ *Methods of life course research: Qualitative and quantitative approaches* / Ed. by J.Z. Giele, G.H. Elder. L: Sage Publications, 1998.

¹⁶ Толстова Ю.Н. Анализ социологических данных: Методология, дескриптивная статистика, изучение связей между номинальными признаками. М.: Научный мир, 2000.

¹⁷ Merton R. The Sociology of Science. Theoretical and Empirical Investigations / Ed. By N. Storer. Chicago: The University of Chicago Press, 1973.

Впоследствии мертоновская модель научного этоса неоднократно дополнялась, а также подвергалась критике, но исходное положение о внутренних правилах научной деятельности оставалось неизменным. Позже получило свое развитие и представление об аномалиях в воспроизведстве научного знания («патологии науки»), отправной точкой которого также послужили работы Р. Мертона.

В ходе анализа институциональных причин, порождающих неполноту использования информационного ресурса и слабое использование вторичных данных, был сделан вывод о наличии следующих симптомов аномального состояния (кризиса) научного сообщества и социологии как научной дисциплины:

- фрагментированность социологического гознания;
- аномалии в воспроизводстве социологического гознания;
- неготовность к переходу на стандарты систематизации;
- коммерциализация и политизация социальной науки.

Граница между нормальным и аномальным состоянием научной дисциплины достаточно операциональна: «нормальное состояние... предполагает достаточно высокую степень согласия ученых относительно проблем, методов, предмета исследования, а также норм поведения - всего того, что делает человека науки предсказуемым и налагает на него определенную дисциплинарную ответственность».¹⁸ С этой точки зрения, любые колебания в сторону снижения степени согласованности надметодологических и методологических установок уменьшают и эвристический потенциал научной дисциплины, подтачивают саму систему социальной организации научной деятельности, порождая *недоверие к результатам предшественников и уменьшая возможность для коллективной работы*. Как следствие, уменьшается и число обращений к вторичным данным. Это усугубляется раздробленностью научного сообщества российских обществоведов, утратой преемственности, оттоком из научных учреждений квалифицированных кадров при минимальном притоке молодежи,¹⁹

¹⁸ Батыгин Г.С., Девятко И.Ф. Миф о качественной социологии // Социологический журнал. 1994. № 2. С. 36. См. также Кун Т. Структура научных революций. М.: Прогресс, 1977.

¹⁹ К 2000 г., по данным КГДежиной, доля исследователей старше 60 лет достигла 20,7%, что намного превышает показатели других стран (Дежина И.Г. Молодежь в науке // Социологический журнал. 2003. № 1).

негативными изменениями в структуре мотиваций и ориентации ученых и преподавателей. В частности, ориентация на коммерческий успех (как одно из последствий функционального кризиса) привела к вмешательству во внутреннюю структуру научных процессов, в результате выбор проблематики исследований перестает зависеть от логики развития научного знания.

Идея цикличности в жизни научного направления получила свое развитие в работе Т.Куна «Структура научных революций» и оказалась весьма плодотворной.²⁰

Согласно куновской концепции, расширенной и дополненной последователями, каждое новое научное направление обычно проходит несколько этапов в своем развитии. На начальном этапе развития научного направления преобладают теоретические разработки, на втором этапе происходит активный сбор и использование первичных данных, появляются первые значимые для развития теории результаты. На следующих этапах возникает необходимость в систематизации накопленного знания и в переосмыслинении базовой теории. Опираясь на описанную логику, использование анализа вторичных данных означает переход на более высокую ступень развития, как исследовательской культуры, так и научного направления в целом. Это предположение было подтверждено серией эмпирических исследований рабочей группы Н. Маллинза.²¹

В параграфе 1.3, «*Эмпиризм versus релятивизм: двумodeli познания в свете проблемы использования данных*», анализируются методологические предпосылки для неполного использования исследовательских

²⁰ Однако ряд ученых (Л. Альтюссер, Г. Вагнер, Н. Луман) подвергают куновскую модель структурных кризисов острой критике. Основной критический постулат заключается в том, что нормальное, согласованное состояние науки является скорее штилем, т.е. непродуктивным периодом, в то время как дискуссии о предмете и методе дисциплины способствуют ее развитию (см., например, Альтюссер Л. Ленин и философия // Пер. с франц. Н. Кулиш. Перевод выполнен по изданию: Louis Althusser. *Lenine et la philosophie suivi de Marx et Lenine devant Hegel*. Ed. Francois Maspero, Paris, 1972 / СПб.: Издательство «Ад Маргинем», 2005; Вагнер Г. Социология: к вопросу о единстве дисциплины // Теория общества: фундаментальные проблемы / Под ред. А.Ф. Филиппова. М.: Канон-Пресс-Ц, Кучково поле, 1999. С. 236-260; Луман К «Что происходит?» и «Что за этим стоит?». Две социологии и теория общества//Современная немецкая социология: 1990-е годы /Под ред. В.В. Козловского, Э. Ланге, Х. Харбаха. СПб.: Социологическое общество им. В.В. Ковалевского, 2002. С. 27-51).

²¹ Маллинз Н. Модель развития теоретических групп в социологии // Социология науки / Составитель Э.М. Мирский. М.: Институт системного анализа РАН, 2000.

данных, связанные со стратегическим разделением на «качественный» и «количественный» подход.

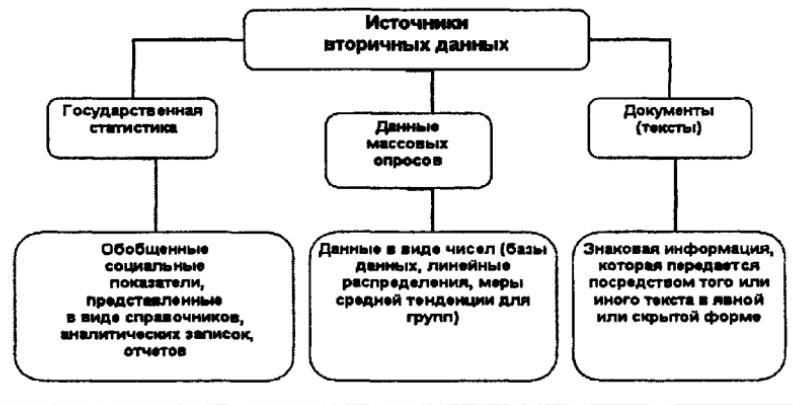
С самого начала своего пути, пишет Н. Луман, «социология приближалась к своему предмету двумя путями: позитивистски и критически».²² Неполное использование данных в количественных исследованиях логическим образом проистекает из проблем прародителя-эмпиризма. Приверженность эмпиризму как традиции опытов и экспериментов породила установку скорее провести свое, новое исследование, чем воспользоваться данными предшественников. Этому сопутствовало превалирование описательной стратегии анализа данных, практика исследований под условным названием «рассмотрим на примере» (практика социального портрета). Неполное использование исследовательского информационного ресурса в качественных исследованиях связано с постулированием «的独特性 (уникальности) случая» и вовлечением в анализ больших объемов описательных данных, а также с запретом доступа к полным текстам по этическим и иным соображениям. В то же время, стратегия качественного исследования чаще всего строится по типу «воронки» (первичное – максимально широкое – постулирование темы, и постепенное отбрасывание ненужного). На выходе исследователь занимается структурированием описательных данных, поэтому момент «излишней» информации уже заложен по определению.

Во второй главе, «*Вторичные данные как информационный ресурс*», проблема неполного использования исследовательских материалов раскрывается с позиции методологии анализа данных в социологии. Анализируются достоинства и недостатки основных источников вторичных данных, показана роль анализа и интерпретации эмпирических данных, среди прочих этапов исследовательского процесса. Делается вывод о том, что выбор тех или иных методов анализа определяется, исходя не только из характера данных (количественные базы данных, транскрипты и т.д.), но также из предполагаемой формы конечного «продукта» исследовательской работы (отчета, научной статьи, учебника).

²² Луман Н. «Что происходит?» и «Что за этим стоит?». Две социологии и теория общества // Современная немецкая социология: 1990-е годы / Подред. В.В. Козловского, Э.Ланге, Х.Харбаха. СПб.: Социологическое общество им. В.В. Ковалевского, 2002. С. 41.

В параграфе 2.1., «Источники вторичных данных: государственная статистика, результаты предшественников, документы (тексты)» рассматриваются различные виды вторичных данных по типу источника, анализируются достоинства и недостатки каждого из них.

Рисунок 2. Источники вторичных данных.



Основным достоинством государственной статистики как источника вторичных данных является возможность работать с обобщенными социальными показателями по стране в целом, по регионам, отраслям, социальным группам. Недостатками считаются невысокая надежность статистических данных, запаздывание информации по срокам. Из этого следует, что статистические данные в качестве вторичных источников могут применяться ограниченно, их следует дополнять перекрестными данными из массовых или экспертных опросов, соответствующих документов и т.д.

Использование результатов массовых опросов (данные предшественников) в качестве вторичного источника подразумевает работу с данными, выраженными в цифровой форме и полученными в исследовательских фирмах, академических институтах, коммерческих организациях. Как вторичный источник, данные предшественников предпочтительны, когда требуются регулярные описания динамики социально-экономических, социально-политических и других показателей, которые на протяжении длительного времени интерпретируются и переводятся на операциональный уровень по одним и тем же типичным схемам у большинства исследователей. Удобны они и для использования в учебных целях. Недостатки связаны с возможной рассогласованностью выборочной и

генеральной совокупности (смещение группообразующих критериев); с недостоверностью тех или иных массовых опросов.

Основными характеристиками такой группы вторичных источников, как документы (тексты) является их *знаковая природа* (принадлежность к специфической знаковой системе), и наличие определенной *информации*, которая передается посредством данного текста в явной или скрытой форме.²³ Анализ литературы позволил сделать вывод о том, что документы и тексты являются наиболее перспективным источником данных с точки зрения повторного использования. Текстовая информация дает уникальную возможность многократного обращения к ней и поэтапного решения тех или иных задач. Богатый интерпретативный ресурс документов и текстов особенно эффективен, когда в фокусе исследования - «история» одного социального объекта, отдельной личности или малой группы, мотивы, поступки и приписываемые этим поступкам смыслы, когда особую важность имеет исторический контекст социальных действий.

В параграфе 2.2., «*Анализ вторичных данных: особенности и приоритетные направления*», на примерах раскрываются содержательные и методические возможности вторичного анализа. Делается вывод о том, что использование вторичных данных является логичным и закономерным при проведении сравнительных, многоцелевых, междисциплинарных исследований.

Эффективное использование вторичных данных подразумевает тщательное изучение выбранных массивов с целью оценки их релевантности, полноты, надежности.²⁴ Это помогает найти и по возможности скорректировать ошибки предшественников.

В работе было выявлено, что на текущий момент наиболее развит анализ вторичных данных в социально-исторических (историческая социология, традиция «устной истории») и социально-демографических исследованиях. ^

²³ Кузнецов В. Герменевтика и ее путь от конкретной методики до философского направления // Логос, 1999. № 10. С. 43-44.

²⁴ Сагането Г.И. Компоненты надежности социологического исследования и проблема сравнимости//Сравнительный анализ и качество социологических данных. М.: ИСИАН СССР, 1984.

²⁵ В практике одного из социально-исторических направлений - устной истории (Р.Грил, П.Хаттон, А.Портелли) - особо подчеркивается идея хранения и ретроспективного обращения к исследовательским материалам.

В параграфе 2.3., «*Data mining: работа с большими массивами вторичных данных*», анализируется опыт узкоспециализированного исследовательского направления по «извлечению данных».

Поскольку с увеличением массивов данных растет их недоиспользованность, методология извлечения данных напрямую служит преодолению этой недоиспользованности.

Быстрое развитие указанного направления в зарубежной исследовательской практике обусловлено высокой актуальностью задач, для решения которых предназначена методология data mining: эффективное использование больших массивов разнограновых вторичных данных. В определенный момент накопленные информационные ресурсы стали слишком большими для реализации традиционных методов анализа. На значительных объемах неоднородных данных (массивы лонгитюдных и кросскультурных исследований, временные ряды, и т.д.) обычные приемы систематизации данных и статистические процедуры не всегда эффективны. Цель направления Data Mining заключается в проработке методологии по обращению с подобными данными и в ее практическом применении, то есть, в выявлении скрытых правил и закономерностей в наборах таких данных.

В третьей главе, «*Операционализация и измерение неполного использования*» предлагается перевод концепции неполного использования социологических данных на операциональный уровень, посредством разработки и применения индекса неполного использования.

В параграфе 3А., «*Индекс неполного использования социологических данных*» отбираются индикаторы (критерии неполного использования) и разрабатывается указанный индекс.

Опираясь на концепцию информационного ресурса, были выбраны следующие критерии: предоставление результатов своих исследований в специализированные архивы, обращение к данным предшественников/использование архивов, проведение сравнительного анализа, проведение вторичного анализа, использование сложных процедур анализа. Эти критерии непосредственно сопряжены с основными институционально-науковед-

²⁶ Backhouse R., Morgan A/. Introduction: is data mining a methodological problem? // Journal of Economic Methodology. 2000. N 7. Vol. 2, P. 171-181.

ческими индикаторами: атомизация / согласованность научного направления, наличие/отсутствие преемственности знания, стадия научной работы (этап накопления или систематизации).

- • Индекс неполного использования подсчитывался суммарно, т.е. каждому критерию присваивался одинаковый вес в 1 балл (положительный в случае отсутствия критерия в исследуемой работе, и отрицательный - в случае наличия критерия). Если критерий нельзя было оценить однозначно, присваивался нулевой балл. Таким образом, максимальный балл для неполного использования информационного ресурса равнялся 5, для полного использования, соответственно, - 5.

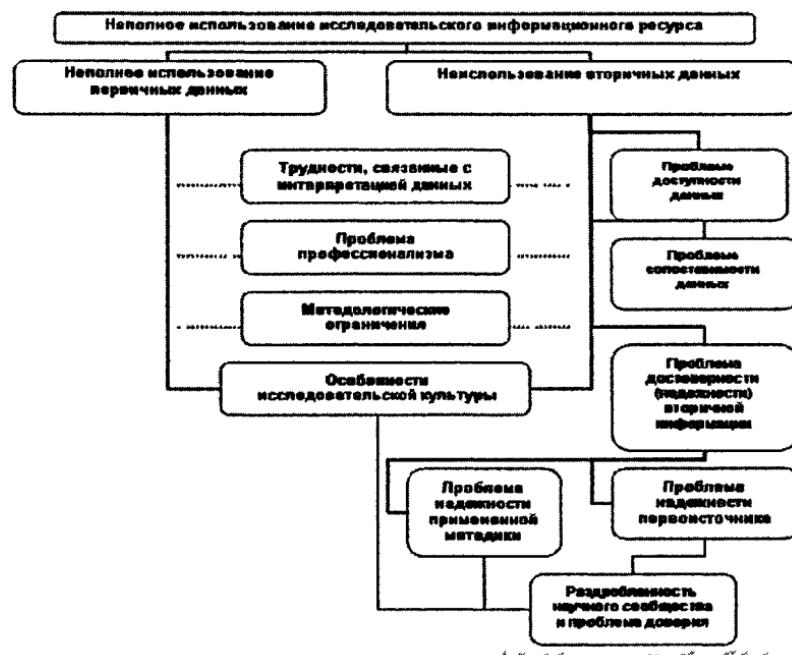
Индекс был подсчитан и проанализирован на публикационном массиве журналов «Социологические исследования», «International Journal of Social Research Methodology», «Oral history». Результаты позволили не только сделать вывод о том, что в российских исследованиях данные нерабатывают свой ресурс ($Mo = 3$, в то время как для проанализированных западных исследований $Mo = -2$.) но и показали, какие именно критерии в наибольшей мере определяют неполноту использования социологических данных: дисбаланс этапов «до-поля» и «после-поля» (89%), отсутствие практики распространения данных и отсылки к архиву (96%).

В параграфе 3.2, «Соотношение до-эмпирических и пост-эмпирических процедур как методическая проблема» анализируется выявленный дисбаланс этапов «до-поля» и «после-поля» как одно из методических препятствий для неоднократного использования данных. Считается, что определенные ограничения на повторное использование данных закладываются самим инструментарием исследования. Однако за тем, что обычно принимают за *влияние метода*, скрывается, на наш взгляд, латентный фактор. Этот фактор мы бы назвали перекосом в сторону разработки «доэмпирических» (теоретико-методологических) процедур, в ущерб процедурам анализа и совершенствования способов предоставления результатов («пост-эмпирических» процедур). Или, на более высоком уровне абстракции, это перекос в соотношении социологической методологии и методологии анализа данных (не в пользу последней). Это приводит к тому, что работа становится описательной, причем вне зависимости от исходной стратегии и характера полученных данных.

В четвертой главе, «*Расширение возможностей для использования*

информационного ресурса в социологии» предлагается классификация причин для неполного использования исследовательского информационного ресурса (см. Рис. 3) и показаны основные направления расширения сферы его использования: развитие систем исследовательских архивов, изменение научной политики, активизация межнаучных коммуникаций, повышение уровня исследовательской культуры. Также резюмируется проработка проблемной ситуации в целом.

Рисунок 3. Дерево проблем неполного использования.



В параграфе 4.1., «*Создание и использование архивов исследовательских данных: зарубежный опыт и российская практика*», последовательно рассматривается зарубежный и российский опыт функционирования различных типов хранилищ: архивов исследовательских организаций, национальных архивов, публикаций первичных материалов, архивов научных и образовательных учреждений, ресурсных центров, по таким критериям, как доступность, поле деятельности, формат данных и др.

Практика ведения количественных баз данных (например, в форматах статистических программ SPSS, SAS, Statistica и других) существует в России на протяжении последних двадцати лет. Однако на регулярной основе поддерживают эту практику только несколько крупных

исследовательских центров: Фонд Общественного Мнения (ФОМ), ВЦИОМ, Левада-Центр (ВЦИОМ-А), РОМИР-Мониторинг, исследовательская группа ЦИРКОН, и некоторые другие. При этом порядок предоставления данных заинтересованным исследователям не отработан и осуществляется ограниченно, либо на коммерческой основе, либо по личным контактам. Это идет вразрез с практикой западных исследовательских организаций, которые, напротив, стремятся к максимальной доступности и расширению круга пользователей своих исследовательских материалов.²⁶

Примером поддержания архивов в академических учреждениях в России является Институт Социологии РАН. Он обладает исключительным правом на распространение компьютерной базы данных «Интегрированная информационная система социальных данных» (INSYS). Еще один архивный проект ИС РАН связан с работой исследовательской группы «Банк данных социологических исследований», сокращённо БДСИ. Исследования в БДСИ разделены на три группы доступа (полный доступ для всех пользователей, доступ для учебно-преподавательских целей и эксклюзивный доступ). Ожидается, что результатом проекта станет не только развитие банка данных, но и внедрение специальных компьютерных программ для структурирования электронных ресурсов и целевого поиска в массивах исследовательских материалов.²⁷

Несмотря на огромное число эмпирических исследований по социологии и смежным дисциплинам, в России нет национального хранилища для подобных материалов. В 2001 г. первые шаги в этом направлении стал осуществлять ВЦИОМ совместно с Независимым институтом социальной политики, при финансовой поддержке Фонда Форда. Так был организован «Единый архив социологических данных», который пока содержит данные примерно двухсот исследований, проведенных до 2001 года. В диссертационной работе делается вывод: чтобы проект «Единый архив социологических данных» вырос до национального архива, необходимо проработать механизм регулярных пополнений архива всеми агентами иссле-

²⁶ См., например, ресурсную базу Центра Э. Роупера - The Roper Center For Public Opinion Research.

²⁷ Ростегаева Н.И. Банк данных социологических исследований: приглашение к сотрудничеству // Социология 4М. 2000. № 12; Ростегаева НИ Банк социологических данных: проблемы функционирования и тенденции развития // Социология 4М. 1998. № 10.

довательского рынка. Также следует обеспечить информационную поддержку проекта, как в России, так и за ее пределами. Хорошо понимая важность информационной поддержки, представители Единого архива начинают принимать участие в тематических конгрессах, конференциях, семинарах с презентациями своего проекта.²⁸

Сравнительно новой российской практикой является создание виртуальных ресурсных центров, число которых за последние 2-3 года постоянно возрастает. Наличие подобных структурированных хранилищ существенно ускоряет поиск информационных ресурсов.²⁹

В параграфе делается вывод, что в России пока еще не сложилась практика обмена «сырыми» данными эмпирических исследований. Предоставление баз данных (и тем более - полнотекстовых транскриптов интервью или фокус-групп) является у нас большой редкостью. Эта ниша еще ожидает своего развития.

В параграфе 4.2, «*Новые направления научной политики и их возможное влияние на эффективность использования социологических данных*», анализируются основные направления текущей реформы научной и образовательной сферы, и оценивается их влияние на эффективность использования исследовательского информационного ресурса. Делается вывод о том, что для уменьшения экспортного дисбаланса в среде научных разработок требуется налаживание механизмов патентообразования и защиты информационного продукта.

В параграфе 4.3, «*Роль междунаучных коммуникаций и исследовательской культуры в преодолении проблемы неполного использования данных*» рассматривается влияние коммуникационной активности и исследовательской культуры на эффективную работу с данными. Часто поиск данных ограничивается «сетями», к которым принадлежит ученый, причем эти сети не всегда имеют разветвленный характер и обмен между ними не всегда эффективен. Делается вывод, что архивы социологической информации могут не только выполнять функцию хранилищ эмпирических результатов,

²⁸ Так, в 2003 г. Единый архив был представлен на Ежегодном конгрессе IASSIST (Международная ассоциация информационного сервиса и технологий для социальных наук) и на Ежегодном конгрессе американской ассоциации славистов (AAASS).

²⁹ Среди наиболее известных и посещаемых проектов - Аудиториум (www.auditorium.ru'), Соционет (www.socionet.ru'), УИС-Россия (www.cir.ru'), Федеральный образовательный портал «Экономика. Социология. Менеджмент» fecsocman.edu.ru).

но и стать группообразующими центрами исследовательской субкультуры по тем или иным дисциплинарным направлениям.

Согласно гипотезе о «незримом колледже», состояние дел в той или иной исследовательской области эффективно контролируется небольшой группой наиболее продуктивных ученых, которые задают направление для актуальных научных разработок. Роль «незримых колледжей» повышается в кризисные для научной дисциплины периоды. Поэтому одной из задач будущих исследований должно стать выявление незримого колледжа тех ученых, которые будут способствовать расширенному использованию вторичных данных.

В *Заключении* диссертации делаются общие выводы, даны некоторые рекомендации, намечены перспективы дальнейшей разработки проблемы неполного использования социологических данных.

В *Приложении* приводятся сравнительные таблицы с описанием основных характеристик российских и зарубежных исследовательских архивов.

Основные положения и результаты исследования нашли отражение в следующих публикациях:

1) Стрельникова А.В. Интернет для социальных наук: вторичный анализ, архивы исследовательских данных и ресурсные центры // Научный сервис в сети Интернет. М.: Изд-во МГУ, 2004. С. 40-43.0,4 пл.

2) Стрельникова А.В. Изучение городских сообществ: социально-исторический подход // Социальное и культурное пространство города: Материалы конференции молодых ученых (27-28 октября 2004, г.Санкт-Петербург). СПбГУ, 2004. С. 59-62.0,2 пл.

2) Стрельникова А. В. Децентрализация управления научной и образовательной деятельностью: проблема эффективности // Современные технологии в социальном управлении. Белгород: БИГМУ - филиал Российской Академии госслужбы, 2005. С. 138-143.0,4 пл.

4) Стрельникова А.В. Исследовательские архивы: расширение возможностей для вторичного анализа//Социологические исследования, 2005. №1. С. 126-131. 0,7 пл.

5) Стрельникова А.В. Новые направления реформы научно-образовательной сферы // Социология управления: сборник научных трудов. Белгород: БИГМУ - филиал Российской Академии госслужбы, 2005.0,4 пл. (в печати).

Принято к исполнению 14/04/2005
Исполнено 15/04/2005

Заказ № 796
Тираж: 100 экз..

ООО «11-й ФОРМАТ» ИНН 7726330900
Москва, Балаклавский пр-т, 20-2-93
(095)747-64-70
www.autoreferat.ru

1598

19 May 2005