

практика, по прошествии 1,5 часов дискуссии начинает явно наблюдаться усталость респондентов, непосредственно влияющая на полноту и длительность их высказываний. Как только модератор фиксирует этот момент, у него остается еще примерно полчаса, чтобы аудитория не ушла в себя и не стала отделяться совсем односложными ответами.

Подводя некоторый итог, необходимо в очередной раз сказать очевидную истину – пандемия и карантинные ограничения оказали огромное влияние на коммуникацию между людьми. Это сильно почувствовала на себе и исследовательская индустрия, и причем не только та ее часть, которая занимается организацией и проведением фокус-групп. Болезни интервьюеров, усилившаяся боязнь респондентов открывать дверь незнакомым людям, которые просят принять участие в анкетном опросе или интервью – это и многое другое заставляет видоизменяться традиционные методики проведения исследований, искать новые способы получения информации. И самое лучшее, что можно предложить в подобной ситуации – это постоянный обмен мнениями, опытом и наработками между исследователями на тематических площадках, форумах и страницах специализированных изданий.

УДК 372.8

ОЦИФРОВКА ИСТОРИЧЕСКИХ МАТЕРИАЛОВ ИССЛЕДОВАНИЙ СОЦИАЛЬНЫХ НАУК КАК ИСТОЧНИК ДАННЫХ СОВРЕМЕННЫХ ИССЛЕДОВАНИЙ

Р. Муракас

Тартуский университет, Уликоли 18, 50090, Тарту, Эстония
rein.murakas@ut.ee

Способность отражать долгосрочные изменения с течением времени и отслеживать прошлые исследования с точки зрения современных парадигм и контекстов добавляет глубины эмпирическому анализу социальных наук. Источниками для этого могут быть не только сохраненные наборы данных, но и бумажная документация бывших исследовательских групп. Фактический анализ таких материалов требует обширной пошаговой работы по оцифровке, включая подготовку машиночитаемого текста с использованием оптического распознавания символов (OCR) и создание адекватных метаданных. Эти процедуры позволяют значительно расширить сравнительный анализ социальных наук, особенно в плане реконтекстуализации предыдущих результатов и метаанализа большого количества исследований.

Ключевые слова: данные эмпирических исследований; документация по обследованию; оцифровка; метаданные; реконтекстуализация.

DIGITIZATION OF HISTORICAL MATERIALS OF SOCIAL SCIENCE RESEARCH AS A SOURCE OF CONTEMPORARY RESEARCH DATA

R. Murakas

University of Tartu, Ülikooli 18, 50090 Tartu, Estonia
rein.murakas@ut.ee

The ability to reflect long-term changes over time and to monitor past research from the perspective of contemporary paradigms and contexts adds depth to the empirical analysis of the social sciences. Sources for this can be not only preserved data sets but also paper documentation of former research teams. The actual analysis of such materials requires extensive step-by-step digitization work, including the preparation of computer-readable text using Optical Character Recognition (OCR) and the creation of adequate metadata. These procedures allow for a significant extension of the comparative analysis of the social sciences, especially in the key of recontextualizing previous research and meta-analysis of many studies.

Keywords: empirical research data; survey documentation; digitization; metadata; recontextualization.

Наблюдение долгосрочных изменений во времени и изучение результатов предыдущих исследований с точки зрения современных парадигм и контекста добавляет глубины эмпирическому анализу социальных наук [1]. К сожалению, возможности в этой области довольно ограничены: данные и аналитические результаты, представленные в доступных предыдущих публикациях, могут быть несопоставимы или недостаточно освещают тему исследования, но оригинальные исторические данные исследований могут быть недоступны, часто они просто уничтожены. Международное архивирование данных [2] в первую очередь охватывает данные крупномасштабных (международных сопоставительных) обследований; потере же индивидуальных данных также способствовало увеличение проектного характера финансирования.

В Эстонии, благодаря деятельности Эстонского архива данных социальных наук (ESTA) [3], значительная часть данных исследований, проведенных в 1970-х и 1990-х годах, была сохранена, но существуют пробелы как в более ранний, так и в более поздний периоды. В то же время можно было сохранить часть бумажной документации, относящейся к деятельности бывших основных исследовательских групп: отчеты об исследованиях, распечатки результатов анализа мэйнфреймов, анкеты, документы, относящиеся к деятельности исследовательских групп, материалы встреч, служебные записки и т. д.

К сожалению, даже в каталогизированном виде такой материал чрезвычайно сложно обрабатывать: на практике, например, приходилось просматривать тысячи страниц бумажного материала, чтобы получить информацию об одной радиопрограмме.

Реальный анализ таких материалов требует обширных мероприятий по оцифровке, основным уровнем которой является сканирование материалов, позволяющее просматривать их в цифровой (графической) форме вместо бумажных документов, а также деятельности, одним из этапов которой является предоставление описаний данных в соответствии со стандартами документов, и их электронное хранение, например, в репозитории.

Этот шаг упростит обработку документов (таких, как запросы по ключевым словам и временным диапазонам дат), но по-прежнему потребует очень большого количества материала для просмотра. Для реального анализа необходимо преобразовать документы в машиночитаемый текст, который можно обрабатывать с помощью программ обработки качественных данных. Часто основные инструменты OCR (оптического распознавания символов) уже включены в программы сканирования документов, но, к сожалению, для старых документов точность такого распознавания текста чрезвычайно низкая. В этой области необходимо использовать специальные программы (например, *Abby FineReader* [4]) с ручной настройкой и исправлением автоматически полученных результатов.

Последняя процедура чрезвычайно трудозатратна, но ее можно делать и поэтапно: 1) создание цифрового графического материала, если возможно, с автоматическим распознаванием текста; 2) обеспечение материала метаданными (описанием данных); 3) публикация в репозитории; 4) распознавание текста и / или ручное исправление, начиная с документов, которые могут представлять наибольший интерес.

В таком процессе формируется большой корпус текста, в котором в более простой версии можно искать требуемый материал, но могут быть выполнены и более сложные процедуры кодирования, основанные на программном обеспечении качественного анализа, что может привести к очень интересным результатам. Числовые данные в материале предлагают дополнительные возможности для формирования графических или табличных результатов.

Конечно, не весь исторический материал можно использовать в виде цифрового текста. Однако рукописные материалы или графические изображения, например, могут быть очень подходящими иллюстрациями для анализа изменений во времени, и для их анализа также доступны программные инструменты.

В заключение отметим, что хранение, оцифровка, описание и обработка всех видов бумажных исследовательских материалов на разных уровнях позволяет значительно расширить временной и сравнительный анализ социальных наук, особенно в том, что касается реконтекстуализации исторических исследований. Также можно использовать метаанализ большого количества исследований. Вот почему важно обеспечить сохранность таких бумажных материалов в современном научном мире, в значительной степени основанном на проектах. Будущим поколениям ученых должны быть гарантированы лучшие аналитические возможности, основанные на постепенной дальнейшей обработке таких данных, чему способствуют и постоянно развивающиеся репозитории цифровой информации [5].

Библиографические ссылки

1. *Tiikmaa H., Lõhmus M.* Public Media Service and Public Interest in Estonia – Changes in Public Service Media 1990–2010 // *Media and Communication*. 2015. No 3. P. 211–232.
2. См. например <https://www.cessda.eu/>.
3. *Murakas R., Rämmer A.* Empirical Research and Situation of Data Archiving in Estonia // *Social Science Data Archives in Eastern Europe: Results, Potentials, and Prospects of Archival Development* / ed.: B. Hausstein, P. de Guchteneire. Berlin: Fergler Verlag, 2002. P. 177–180.
4. См. <https://pdf.abbyy.com/finereader-pdf/>.
5. См. <https://zenodo.org/>; <https://www.openaire.eu/>.

УДК 004.55

ЦИФРОВАЯ СРЕДА КАК НОВАЯ РЕАЛЬНОСТЬ КОММУНИКАЦИИ

В. И. Привалова

Казанский федеральный университет
Институт филологии и межкультурной коммуникации,
ул. Татарстан, 2, зд. 33, 420021 Республика Татарстан, г. Казань, Россия
ivprivalova@mail.ru

Цифровая среда стала новой реальностью существования коммуникативного пространства. Преобразования в модели коммуникативного процесса стали возможными в связи с появлением Интернета и единой цифровой среды. С формированием пользовательского контента оформилась двухчленная модель коммуникации, где наблюдается совмещение ролей отправителя и получателя наряду с циклическим распространением информации. Другими особенностями