



Социологические науки

УДК 004.65+303.02

А.Ю. Дегтярева

Дегтярева Айжан Юрьевна, бакалавр 4 курса группы ИС-15 ФСКДТ Краснодарского государственного института культуры (Краснодар, ул. им. 40-летия Победы, 33), e-mail: aizhan.degtyareva@gmail.com

Научный руководитель: **Яхина Елена Петровна**, кандидат педагогических наук, доцент кафедры экономики и информационных технологий Краснодарского государственного института культуры (Краснодар, ул. им. 40-летия Победы, 33), e-mail: eyakhina@gmail.com

ОТКРЫТЫЕ БАЗЫ ДАННЫХ СОЦИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ

Статья посвящена открытым базам данных, предметной областью которых являются социологические исследования. Подчеркивается, что открытые данные являются удобным источником информации для различных видов человеческой деятельности.

Ключевые слова: базы данных, форматы организации данных, банки данных, датасеты, социологические исследования.

A.Yu. Degtyareva

Degtyareva Aizhan Yuryevna, bachelor of the 4th course of IS-15 group of the Krasnodar state institute of culture (33, im. 40-letiya Pobedy st., Krasnodar), e-mail: aizhan.degtyareva@gmail.com

Research supervisor: **Yakhina Elena Petrovna**, candidate of pedagogic sciences, assistant professor of department economy and information technologies of the

Krasnodar state institute of culture (33, im. 40-letiya Pobedy St., Krasnodar),
e-mail: eyakhina@gmail.com

OPEN SOCIOLOGICAL RESEARCH DATABASES

The article considers the open databases, the subject area of which is a sociological research. The significance of the open data as a convenient source for different human activities is emphasised.

Key words: database, data organization format, data bank, datasets, sociological researches.

В современном понимании базы данных определяются как совокупность данных, упорядоченных и организованных для обработки с помощью ЭВМ [1]. Область применения баз данных охватывает практически все сферы человеческой деятельности – от бухгалтерского учета для малого бизнеса до масштабных баз данных правительственного уровня. Разновидностью баз данных являются открытые данные – базы данных со свободным доступом для любых пользователей [12].

В классификации баз данных по способу доступа различают локальные базы данных (доступ к данным возможен только на устройствах в пределах локальной сети) и сетевые базы данных (доступ к данным возможен с удаленных устройств) [2]. В отличие от локальных баз данных, сетевые базы данных размещаются на централизованном сервере, доступ к которому пользователи могут получить из любой точки планеты. Открытые данные относятся к сетевым базам данных, информация из которых может скачиваться пользователем напрямую с сайта либо может быть получена в электронном виде в результате официального запроса в организацию.

Наиболее популярные форматы организации открытых данных – это банки данных (архивы) и отдельные датасеты.

Банки данных представляют собой централизованные базы данных, включающие множество отдельных файлов. Такие банки данных, как правило, публикуются официальными источниками, например, институтом социологии РАН. Кроме того, для удобства пользователя данные в централизованных архивах обычно упорядочиваются по каталогам. В таких банках данных публикуются как структурированные данные (таблицы), так и неструктурированные данные (полнотекстовые документы, статьи, сборники и т.д.) [3]. Банки данных и архивы социологических исследований широко используются для научных исследований и в образовательных целях. Этому способствует подтвержденная достоверность данных и авторитетность источника.

Датасеты представляют собой отдельные таблицы больших объемов [6]. В отличие от банков данных, датасеты создаются как организациями, так и индивидуальными пользователями и могут быть размещены как на специализированных сайтах, так и в любом другом месте в Интернете. Поиск необходимых датасетов зачастую затруднен из-за большого разброса данных по разным источникам, однако существуют специальные поисковики, которые успешно решают данную проблему.

Датасеты, содержащие социологические данные, находят свое применение в коммерческих целях. На их основе проводится статистическая аналитика, выводятся различные закономерности и строятся прогнозы. Кроме того, датасеты являются основным источником данных для машинного обучения и создания искусственного интеллекта. Причиной этому являются: большой объем данных, содержащихся в таблицах, и формат, удобный для обработки с помощью алгоритма [11].

Существует огромное количество банков данных и архивов социологических исследований. Например, большой популярностью пользуется банк социологических данных института социологии Российской

академии наук. В банке содержится более 700 отдельных исследований за период с 1966 года, проведенных самим институтом социологии РАН либо другими социологическими центрами [3].

Другим известным источником социологических исследований является Всероссийский центр изучения общественного мнения (ВЦИОМ), который проводит исследования в области социологии, политики и маркетинга с 1987 года. В базе данных ВЦИОМ материалы делятся по тематическим разделам и проектам, а для удобства пользователя реализован поиск по материалам с использованием фильтров [7].

Крупные учебные заведения также могут иметь собственные архивы данных. Например, существует единый архив экономических и социологических данных, созданный научно-исследовательским институтом Высшей школы экономики. В базе можно найти результаты различных опросов, таблицы, статистические ряды и многое другое [8].

В отличие от банков данных и архивов, датасеты редко публикуются одним централизованным источником. Следовательно, для поиска датасетов существуют специализированные поисковики и тематические сайты. Примером подобного сайта является Enigma Public – портал, собирающий в единую базу данных тысячи датасетов из различных источников. Enigma Public представляет пользователю поиск по тематикам, тегам, описание датасета и ключевые слова, а также возможность скачать датасет для свободного пользования [9].

В сентябре 2018 года компания Google запустила специальный поисковик для датасетов – Dataset Search. Этот сервис индексирует датасеты по всему интернету с помощью специальных дескрипторов и выдает пользователю ссылки на страницы, на которых эти датасеты размещены. Кроме того, Dataset Search поддерживает поиск на различных языках. До сих пор Dataset Search существует только в бета-версии [5; 10].

Подводя итог, можно сделать вывод о том, что открытые данные социологических исследований являются ценным и удобным источником

информации для различных видов деятельности человека – образования, журналистики, коммерции, науки. Принцип открытости данных повышает требования к их качеству и уровень ответственности исследователей за достоверность информации, а также способствует развитию демократического общества в целом [4].

Список используемой литературы:

1. База данных. URL: https://ru.wikipedia.org/wiki/База_данных (дата обращения 1.04.2019)
2. Классификация баз данных. URL: <https://studopedia.info/8-79275.html> (дата обращения 1.04.2019)
3. Банк социологических данных. URL: <http://www.isras.ru/Databank.html> (дата обращения 1.04.2019)
4. Открытость социологических данных повышает стандарты исследовательской работы. URL: http://www.sociologos.ru/cennosti_i_standarty/Otkrytost_sociologicheskikh_dannyh_povyshaet_standarty_issledovatelskoj_raboty (дата обращения 1.04.2019)
5. Google Dataset Search. URL: <https://toolbox.google.com/datasetsearch> (дата обращения 1.04.2019)
6. Data set. URL: https://en.wikipedia.org/wiki/Data_set (дата обращения 1.04.2019)
7. База социологических данных ВЦИОМ. URL: <https://wciom.ru/database/> (дата обращения 1.04.2019)
8. Единый архив экономических и социологических данных. Режим URL: <http://sophist.hse.ru> (дата обращения 1.04.2019)
9. Enigma Public is the world's broadest collection of public data. URL: <https://public.enigma.com/about> (дата обращения 1.04.2019)

10. Making it easier to discover datasets. URL: <https://www.blog.google/products/search/making-it-easier-discover-datasets/> (дата обращения 1.04.2019)

11. Training, Validation and Test Sets. URL: https://en.wikipedia.org/wiki/Training,_validation,_and_test_sets (дата обращения 1.04.2019)

12. Открытые данные. URL: https://ru.wikipedia.org/wiki/%D0%9E%D1%82%D0%BA%D1%80%D1%8B%D1%82%D1%8B%D0%B5_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D0%B5 (дата обращения 1.04.2019)