

Научная статья

УДК 911.3:33+004.42+004.6

doi: 10.17072/2079-7877-2024-4-54-66

**ПРИМЕНЕНИЕ ОТКРЫТЫХ ДАННЫХ НАЛОГОВОЙ СЛУЖБЫ
ДЛЯ ЭКОНОМИКО-ГЕОГРАФИЧЕСКОГО АНАЛИЗА****Павел Олегович Сёмин**

Пермский государственный национальный исследовательский университет, г. Пермь, Россия

Аннотация. В статье представлена технология создания геокодированного табличного набора данных о субъектах малого и среднего предпринимательства (МСП) в России на основе открытых данных Федеральной налоговой службы (ФНС России). Получаемый в результате применения технологии набор данных охватывает всю территорию России и содержит информацию, детализированную до уровня конкретных юридических лиц и индивидуальных предпринимателей. Таблица включает следующие поля (колонки): ИНН, ОГРН, вид (юридическое лицо, индивидуальный предприниматель, глава крестьянского (фермерского) хозяйства), категория (микропредприятие, малое, среднее предприятие), имя или наименование, место жительства или местонахождение (субъект федерации, район, город, населённый пункт), код основного вида деятельности по ОКВЭД, доходы, расходы, среднесписочная численность работников. Данные о доходах, расходах и численности работников доступны за каждый год с 2018 г., остальные – за каждый месяц с августа 2016 г. Статья излагает воспроизводимую методику обработки исходных данных и демонстрирует её применение для получения и разведочного анализа сведений о фирмах в сфере сельского, лесного хозяйства и рыболовства. Дополнением к статье является открытое консольное приложение на языке программирования Python, реализующее описанную технологию. Показано, что открытые данные ФНС России с помощью предложенного алгоритма можно применять для решения ряда аналитических и исследовательских задач в области экономической географии, в особенности тех, где требуются дезагрегированные сведения или пространственная детализация до населённых пунктов. Благодаря географическим координатам получаемый набор данных может без дополнительной обработки использоваться для картографирования. Наличие кодов ОКТМО позволяет объединять его с официальными статистическими сведениями.

Ключевые слова: открытые данные, пространственный анализ, геоданные, административные данные, малое и среднее предпринимательство, реестр организаций, Apache Spark, ФНС России

Для цитирования: Сёмин П.О. Применение открытых данных налоговой службы для экономико-географического анализа // Географический вестник = Geographical bulletin. 2024. № 4 (71). С. 54–66. doi: 10.17072/2079-7877-2024-4-54-66

Original article

doi: 10.17072/2079-7877-2024-4-54-66

APPLICATION OF THE TAX SERVICE OPEN DATA FOR ANALYSIS IN ECONOMIC GEOGRAPHY**Pavel O. Syomin**

Perm State University, Perm, Russia

Abstract. This paper presents a methodology for the creation of a geocoded tabular dataset of small and medium-sized enterprises (SMEs) in Russia based on open data provided by the Federal Tax Service (FTS) of Russia. The resulting dataset encompasses the entire territory of the country. The data is provided at the level of individual SMEs. The dataset is structured as a CSV file comprising the following fields: tax number, registration number, legal status (juridical person, sole trader, head of a peasant (farm) enterprise), SME category (microbusiness, small-sized business, medium-sized business), name, address (region, district, city, settlement), main activity code according to OKVED (Russian Classifier of Economic Activity Types), income, expenses, and average number of employees. The dataset includes revenue, expenses, and employee data from 2018 onward, with yearly granularity; all the other variables are presented from August 2016 onward, with monthly granularity.

The article presents a reproducible methodology for the processing of raw FTS data and illustrates its application in the generation and exploratory data analysis of a dataset comprising firms in the agriculture, forestry, and fishery sectors. A reference implementation of the described technology is provided in the form of an open-source Python command-line tool. The paper demonstrates that the proposed technique enables the utilization of FTS open data to address a range of analytical and academic tasks in the field of economic geography, particularly those benefiting from disaggregated information or requiring spatial resolution at the settlement level. Furthermore, the incorporation of geographic coordinates into the dataset facilitates direct mapping without additional processing needed. The inclusion of municipal codes allows for seamless integration with official statistical information.

Keywords: open data, spatial analysis, spatial data, administrative data, small and medium-sized enterprises, business registry, Apache Spark, FTS of Russia

For citation: Syomin, P.O. (2024). Application of the tax service open data for analysis in economic geography. *Geographical Bulletin*. No. 4(71). Pp. 54–66. doi: 10.17072/2079-7877-2024-4-54-66

Введение

Современные общественно-географические исследования часто основаны на работе с данными. К таким данным относятся, например, агрегированная (сводная) социально-экономическая статистика [5, 8, 26], общедоступные сведения социальных сетей и поисковых систем [2, 13], геоданные сотовых операторов [3, 4, 9], информация коммерческих систем бизнес-аналитики [14, 18] или геоинформационных сервисов [16], спутниковые снимки [15]. По объёму, разнообразию и скорости обновления некоторые источники информации приближаются к характеристикам «больших данных» [1], использование которых для решения традиционных общественно-географических задач – потенциально перспективное направление [7].

Для части экономико-географических исследований важны сведения, детализированные до уровня отдельных фирм. К таким сведениям относятся, например, местонахождение, выручка, расходы, число работников, наименование и другие характеристики. Указанные данные могут использоваться напрямую, как в работах К.В. Ростислава [23], Е.А. Коломак и А.И. Шерубнёвой [12], которые анализируют агломерационные эффекты на основе показателей предприятий. Кроме того, на их основе можно считать статистику с произвольной пространственной или временной группировкой. Например, И.С. Гуменюк оперирует подсчётами абсолютного и относительного числа малых и средних предприятий по муниципалитетам Калининградской области [10]; М.А. Саранча включает количество предприятий в интегральный индекс уровня развития малого предпринимательства на территории муниципального образования [24]; М.А. Макушин с соавторами использует данные о предприятиях в городах БАМа, чтобы оценить структуру промышленности [17]. Дополнительную важность сведения о конкретных фирмах приобретают при исследованиях в масштабе городов [11, 16, 19], при изучении узких секторов экономики, которые не фигурируют обособленно в статистике [18, 20], при анализе отдельно взятой отрасли [14].

Удобного источника геопривязанной информации о фирмах в России нет. Исследователи и прикладные специалисты обычно опираются на данные коммерческих («2ГИС», «Яндекс-карты») [16] или открытых (OpenStreetMap, Wikimapia) геоинформационных сервисов, систем бизнес-аналитики (например, Spark, «Контур.Фокус» и Ruslana) [14, 18]. Как правило, коммерческие фирмы предоставляют сведения за существенную плату, а для выгрузки и обработки информации открытых ресурсов требуется приложить немало усилий. Формально полный список российских организаций и индивидуальных предпринимателей содержится в двух государственных реестрах (Едином государственном реестре юридических лиц и Едином государственном реестре индивидуальных предпринимателей соответственно), однако ни один из них не распространяется в полном виде – можно лишь извлекать отдельные записи вручную через веб-сервис ФНС России либо автоматизированно за плату через программный интерфейс. Таким образом, актуален вопрос об удобном инструменте, который позволил бы бесплатно и легально получить максимально полный набор геопривязанных данных о российских фирмах.

С недавних пор экономико-географы обратили внимание на такой источник информации, как административные данные ФНС России. Для исследований используются, в частности, сведения о доходах физических лиц [21], о налоговых поступлениях в разрезе субъектов России или видов экономической деятельности [22, 25], о среднесписочной численности работников организаций [23], о количестве учтённых жилых помещений [16], о количестве малых и средних предприятий и мерах их поддержки [10]. Особое внимание привлекает реестр субъектов малого и среднего предпринимательства (реестр МСП), который регулярно и полностью размещается на сайте службы и содержит сведения о конкретных организациях и индивидуальных предпринимателях, включая адреса регистрации. Де-факто это наиболее полный набор данных о российских фирмах, детализированный до уровня отдельных фирм, доступный всем без платы и геопривязанный хотя бы каким-то образом. Использовать его в исходном виде – непростая задача: выгрузки реестра за несколько лет имеют суммарный сжатый объём 200 Гб и характеризуются сложной многоуровневой структурой, а адреса, записанные в привычном для человека виде, не содержат географических координат или кодов муниципальных образований, в результате чего мало пригодны для автоматизированного анализа. Цель работы, проведённой при подготовке данной статьи, – упростить использование данных реестра МСП для широкого круга экономико-географов.

В статье предложена технология (алгоритм), позволяющая формировать геопривязанный табличный набор данных о малых и средних фирмах в России. В качестве основы выбран реестр субъектов малого и среднего предпринимательства (реестр МСП), размещённый в форме открытых данных на сайте Федеральной налоговой службы России (ФНС России). В отношении субъектов малого и среднего предпринимательства (МСП) набор данных представляет собой генеральную совокупность. Кроме того, в статье обосновывается, что указанный набор данных может использоваться как достаточно репрезентативная выборка для анализа экономики в целом, а не только малого и среднего предпринимательства. Статья сочетает в себе элементы методической работы, data rare и software rare: она подробно описывает исходные данные и воспроизводимый алгоритм (методику) работы с ними, в том числе содержит ссылку на код программы, после чего демонстрирует результат обработки данных на тестовом примере, включая потенциальный вариант разведочного анализа.

Материалы и методы

Технология опирается на один основной источник исходных данных и несколько вспомогательных источников. Все они являются общедоступными и бесплатными для использования.

Реестр МСП как основной источник сведений. Реестр МСП – это государственная информационная система, содержащая информацию о субъектах малого и среднего предпринимательства. Она предусмотрена ст. 4.1 Федерального закона от 24.07.2007 № 209-ФЗ «О развитии малого и среднего предпринимательства в Российской Федерации». Субъекты малого и среднего предпринимательства (субъекты МСП) – это коммерческие юридические лица и индивидуальные предприниматели, соответствующие критериям по доходу, числу работников и некоторым другим показателям. За реестр отвечает ФНС России. Сведения в реестр попадают автоматически, то есть без участия самого субъекта МСП, благодаря чему обеспечивается максимально возможная полнота данных. По закону, сведения реестра ежемесячно размещаются в Интернете и являются общедоступными в течение пяти лет с года опубликования. Норма о реестре появилась в законе в конце 2015 г. На практике реестр размещён на сайте ФНС России в форме открытых данных в виде ежемесячных полных выгрузок, первая из которых датируется августом 2016 г., а последняя доступная на момент подготовки статьи к публикации – июлем 2024 г. (<https://www.nalog.gov.ru/opendata/7707329152-rsmp/>).

Технически реестр МСП в форме открытых данных – это коллекция zip-архивов xml-файлов. Каждый zip-архив – это «срез» (выгрузка, дамп, dump) реестра на определённую дату. Коллекция zip-архивов тем самым образует временной ряд (time series). Каждый zip-архив содержит множество (обычно 6–7 тыс.) xml-файлов. Каждый xml-файл включает в себя информацию о множестве (обычно ровно 900) субъектов МСП. Один субъект МСП в xml-файле представлен элементом «Документ». Отдельные характеристики субъекта МСП представлены вложенными элементами, атрибутами и значениями в элементе «Документ». К числу таких характеристик относятся ИНН, ОГРН, адрес регистрации, наименование или имя, коды ОКВЭД, вид и категория субъекта МСП, дата включения в реестр и другие.

На момент написания статьи (декабрь 2023 г.) было доступно 88 zip-архивов. Их суммарный объём – 193 Гб в сжатом виде и 2,4 Тб в распакованном. Из-за большого объёма исходных данных требуется оптимизировать алгоритм их обработки и применять специализированные инструменты для работы с «большими данными», о которых будет сказано ниже. При этом проделанные оптимизации позволяют осуществить все необходимые вычисления на обычном персональном компьютере, не задействуя облачные платформы.

Вспомогательные источники. В дополнение к реестру МСП используются ещё два набора открытых данных ФНС России. Первый – это «Сведения о суммах доходов и расходов по данным бухгалтерской (финансовой) отчетности организаций» (<https://www.nalog.gov.ru/opendata/7707329152-revexp/>). Он содержит ИНН, суммарные доходы и расходы за год, предшествующий году публикации. Обновляется несколько раз в год, но последующие обновления в течение одного года, по-видимому, представляют собой уточнение ранее опубликованных сведений, поэтому практическая периодичность обновления составляет год. В отличие от реестра МСП, содержит только информацию об организациях, но не об индивидуальных предпринимателях. Технически аналогичен реестру МСП: представляет собой временной ряд zip-архивов xml-файлов, где каждый архив соответствует определённой дате, а каждый xml-файл внутри архива содержит информацию о множестве организаций. Второй вспомогательный источник информации – «Сведения о среднесписочной численности работников организации» (<https://www.nalog.gov.ru/opendata/7707329152-sshr2019/>). Содержит ИНН и среднесписочную численность работников за год, предшествующий году публикации сведений. По остальным характеристикам идентичен предыдущему. Данные обоих наборов доступны с 2019 г. по настоящее время. На момент подготовки статьи на сайте ФНС России была опубликована только последняя версия вспомогательных наборов данных. Раньше были доступны предыдущие. Они сохранились в личном архиве автора и доступны по запросу.

Для нормализации адресов, расстановки координат и кодов ОКТМО используются два справочника. Один из них – «Населенные пункты России: численность населения и географические координаты» (<https://data.rcsi.science/data-catalog/datasets/160/>). Подготовлен проектом «Инфраструктура научно-исследовательских данных» на базе открытых данных Росстата и сведений региональных медицинских информационно-аналитических центров. Содержит перечень городских и сельских населённых пунктов с указанием субъекта Российской Федерации и муниципального образования, кода ОКТМО и географических координат. Второй – «Города России» (<https://github.com/hflabs/city>). Подготовлен компанией «Дадата» на базе собственных разработок. Содержит перечень городских населённых пунктов с информацией о их месте в административно-территориальном делении, коде ОКТМО и географических координатах. Поскольку некоторые города в нём отсутствуют, автор статьи добавил их вручную, опираясь на сведения Федеральной информационной адресной системы (ФИАС).

Методика обработки данных. Предложенный алгоритм обработки данных состоит из пяти шагов, некоторые из которых разбиты на подзадачи. Он изображён на схеме (рис. 1) и описан ниже. Обозначения в скобках соответствуют обозначениям на схеме.

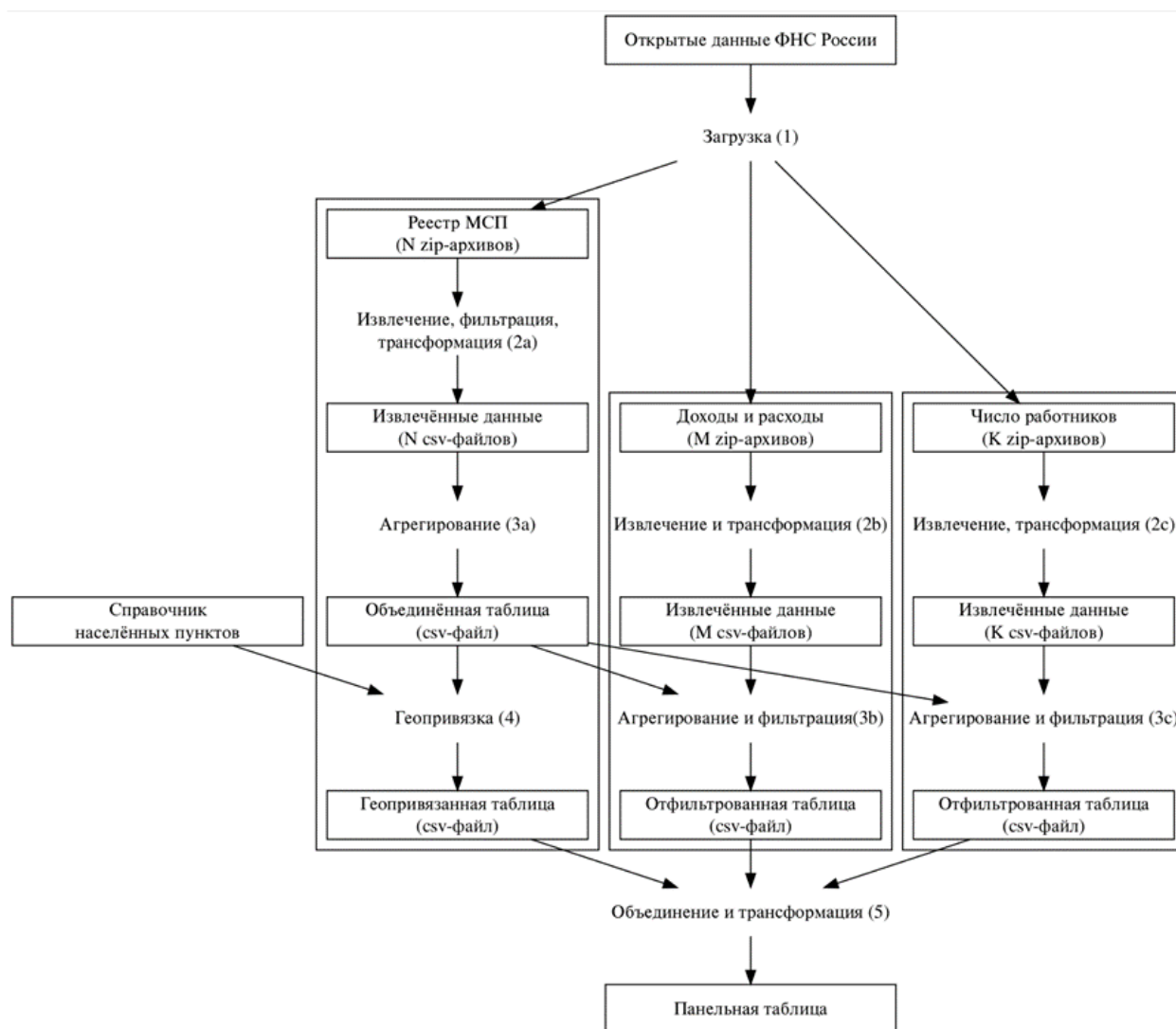


Рис. 1. Схема обработки данных
Fig. 1. Data processing diagram

Первый шаг – это загрузка данных ФНС России (1). Наборы открытых данных загружаются с сайта ФНС России на устройство пользователя. Загрузка ускоряет дальнейшую обработку. Кроме того, благодаря ей создаётся резервная копия исходных данных, что страхует от риска удаления наборов с сайта налоговой службы. Гео-ретически также возможна загрузка и обработка данных в облачном сервисе, таком как Yandex Cloud.

Второй шаг – фильтрующее извлечение данных (2а). Каждый zip-архив реестра МСП преобразуется в csv-таблицу. Каждая строка таблицы содержит сведения об одном субъекте МСП. Извлекаются следующие характеристики: вид, категория, ИНН, ОГРН, имя или наименование, дата включения в реестр, дата выгрузки сведений в набор открытых данных, юридический адрес (с точностью до населённого пункта), основной код ОКВЭД. При извлечении данные могут фильтроваться по одному или нескольким кодам или группе кодов ОКВЭД. Отфильтрованная таблица будет содержать сведения только из необходимой части реестра МСП, что ускоряет работу и сокращает использование дискового пространства. Реализация данного шага оптимизирована, чтобы уменьшить время обработки, расход оперативной памяти и места для хранения временных файлов: данные из архивов читаются напрямую без распаковки архива целиком, при этом извлекаются и сохраняются только необходимые атрибуты записей о фирмах, а всё лишнее отбрасывается. Аналогичным образом обрабатываются сведения о доходах, расходах (2b) и о среднесписочной численности работников организаций (2c).

Третий шаг – агрегация данных о субъектах МСП (3а). Таблицы в csv-формате, получившиеся на шаге 2а, сводятся в одну таблицу, из которой удаляются дубли, то есть строки, в которых совпадает набор значимых признаков: ИНН, ОГРН, имя или наименование, вид, категория, код ОКВЭД, адрес регистрации. После удаления дублей в каждой строке проставляются две даты: дата начала действия и дата окончания действия. Эти даты указывают на период, в течение которого актуальна информация, содержащаяся в строке. Например, если организация с 10 августа 2016 г. по 10 марта 2021 г. была известна под одним наименованием, а с 10 апреля 2021 г. –

под другим, то в агрегированной таблице она будет представлена двумя строками, у первой из которых дата начала – 10.08.2016, дата окончания – 10.03.2021; у второй дата начала – 10.04.2021, дата окончания – 10.12.2023 (условная дата последнего архива исходных данных). Кроме того, на данном этапе происходят вспомогательные преобразования, цель которых – исправить неточности в исходных данных и корректно определить дубли. Например, в реестре МСП тип населённого пункта может быть указан полностью («город») или сокращённо («г.»), и вспомогательная обработка перед агрегацией позволяет игнорировать такие различия и воспринимать адрес как неизменный. Аналогичным образом производится агрегация данных о доходах, расходах (3b) и агрегация данных о работниках (3c). Кроме того, эти данные не только агрегируются, но и фильтруются так, чтобы в них осталась только информация о тех организациях, которые есть в основном наборе, полученном на шаге 3а.

Четвёртый шаг – геопривязка субъектов МСП (4). Каждая запись в csv-таблице, полученной на этапе 3а, дополняется кодом ОКТМО и географическими координатами населённого пункта, в котором находится субъект МСП. Для этого используются адрес регистрации и вспомогательные наборы геоданных. Кроме того, исходный адрес в наборе данных заменяется на нормализованный. После нормализации адресов набор данных повторно проверяется на дубли. Дубли, образовавшиеся из-за нормализации адресов, удаляются, а даты актуальности сведений пересчитываются.

Пятый, опциональный шаг – создание панельного представления (5). Таблица, полученная на этапе 4, объединяется с таблицами, полученными на этапах 3b и 3c, и преобразуется в панельный временной ряд, где представлены характеристики каждого субъекта МСП за каждый год с 2016. Содержательно такая таблица не отличается от первоначальных, но более удобна для анализа: она одна вместо трёх, и вместо компактного, но не слишком интуитивного и весьма технологизированного формата с датами актуальности информации в ней представлена по годам. Если сведения о субъекте МСП в течение года менялись, то строк будет несколько. Такой вид более привычен для исследователей, а также упрощает объединение набора данных с официальной статистикой, которая также обычно даётся в разрезе календарных лет.

Характеристики итогового набора данных. Получаемый набор данных охватывает всю территорию России за период с августа 2016 г. по настоящее время. В зависимости от настроек фильтрации он содержит информацию обо всех субъектах МСП или только о тех, которые ведут деятельность в выбранных сферах. Его пространственное разрешение – до конкретного населённого пункта, временное – до месяца, структурное – до конкретного кода по общероссийскому классификатору видов экономической деятельности (ОКВЭД). Сведения являются дезагрегированными, то есть детализированными до уровня отдельных фирм (организаций и индивидуальных предпринимателей). Помимо структурированных адресов, набор содержит географические координаты населённых пунктов, в которых находятся фирмы, и коды ОКТМО соответствующих муниципальных образований.

Технические детали реализации. Описанный выше алгоритм реализован как консольное (CLI) приложение на языке программирования Python. Чтение xml осуществляется с помощью высокоэффективной библиотеки lxml. Для агрегации, геопривязки и создания панельного представления используются система обработки больших данных Apache Spark и библиотека Pandas. Исходный код опубликован в личном репозитории автора статьи: <https://github.com/PavelSyomin/ru-smb-companies>.

Демонстрационный вариант. С помощью описанного алгоритма был подготовлен пробный набор данных о субъектах МСП, ведущих деятельность в сфере сельского, лесного хозяйства, охоты, рыболовства и рыбноводства. Фильтрация вида деятельности производилась по следующему условию: основной код ОКВЭД организации или предпринимателя находится в группе А классификатора. Использовались исходные данные за весь доступный на момент написания статьи период: с августа 2016 г. для реестра МСП; с октября 2019 г. для сведений о доходах и расходах организаций; с марта 2020 г. для сведений о среднесписочной численности работников организаций – и до ноября 2023 г. Описание и визуализация результата приведены в следующем разделе статьи.

Результаты и обсуждение

Техническая статистика работы алгоритма. Время создания демонстрационного набора данных – около 15 часов. Больше всего занимает извлечение информации из архивов (шаг 2). Объём исходных данных – около 200 Гб. Объём итоговых таблиц – 60 Мб, панельного представления – 634 Мб. Итоговый набор данных содержит информацию примерно о 102 тыс. организаций и 316 тыс. индивидуальных предпринимателей.

Распределение по субъектам Российской Федерации. Нормализованные адреса позволяют агрегировать информацию по субъектам Российской Федерации. На рис. 2 представлено количество субъектов МСП, которые вели деятельность в сфере лесного (слева) и сельского (справа) хозяйства в 2021 г. Распределение ожидаемо разное: наибольшее количество фирм в сфере сельского хозяйства расположено в Черноземье и на юге России, также заметны животноводческие регионы; лесное хозяйство сосредоточено на севере европейской части страны, в Восточной Сибири и на Дальнем Востоке. Кроме того, видно, что по числу зарегистрированных фирм лесное хозяйство уступает сельскому: даже в регионах, которые специализируются на лесном хозяйстве, количество субъектов МСП, ведущих основную деятельность в этой сфере, едва сравняется с количеством субъектов МСП, ведущих деятельность в сельском хозяйстве (вероятно, в основном в животноводстве и охоте, которая по ОКВЭД отнесена к тому же коду 01, что и сельское хозяйство).

Экономическая, социальная и политическая география
Сёмин П.О.

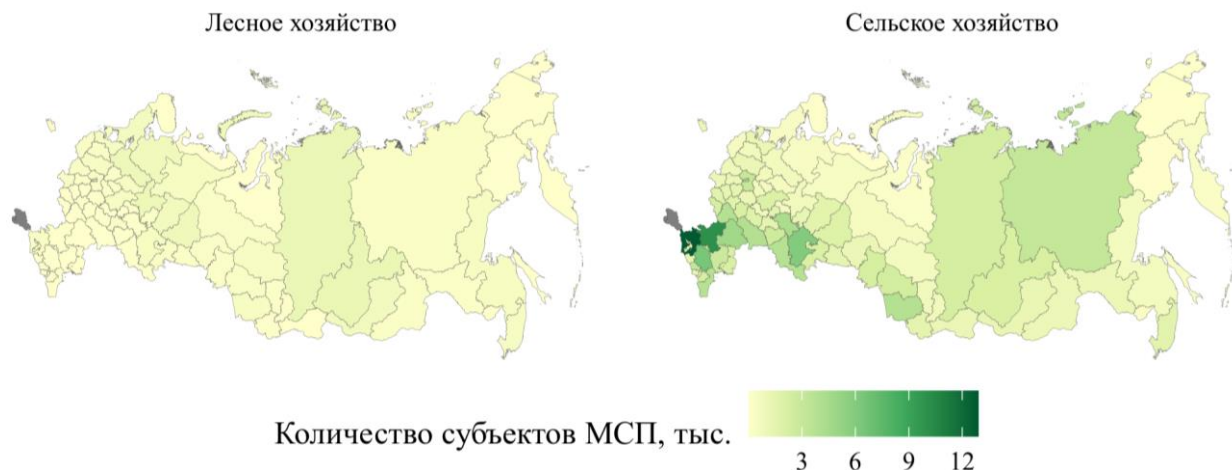


Рис. 2. Распределение количества субъектов МСП, ведущих деятельность в сфере сельского и лесного хозяйства, по регионам (субъектам Российской Федерации) в 2021 г.

Fig. 2. Regional distribution of SMEs with main activity in agriculture or forestry, 2021

В отличие от официальной статистики, производимый набор данных позволяет проанализировать пространственное распределение с точностью до конкретных видов деятельности, а не крупных агломератов. На рис. 3 показано количество субъектов МСП, ведущих деятельность по выращиванию пшеницы (код 01.11.11) и риса (код 01.12) в 2021 г. Можно заметить, что пространственное распределение фирм отличается и отчасти согласуется с районами выращивания соответствующих культур, хотя вряд ли полностью соответствует ему, так как номинальный код основного вида деятельности не всегда соответствует фактическому. Кроме того, выращивание пшеницы – это намного более распространённое занятие, чем выращивание риса, где даже в регионах специализации общее число фирм весьма невелико.

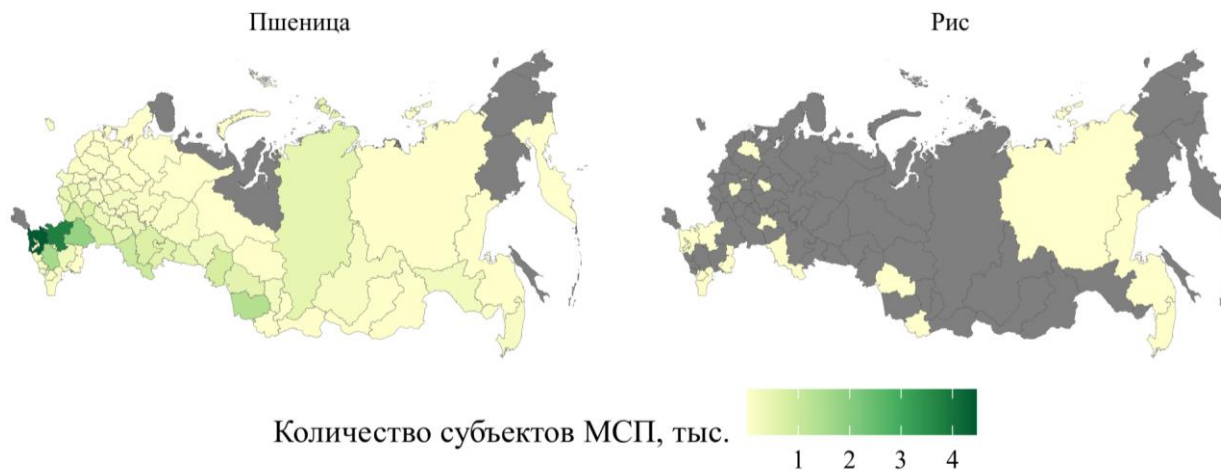


Рис. 3. Распределение количества субъектов МСП, ведущих деятельность по выращиванию пшеницы и риса, по регионам (субъектам Российской Федерации). Серый цвет – отсутствие данных

Fig. 3. Regional distribution of SMEs whose main activity code is growing wheat or rice, 2021.

Grey indicates no data available

Распределение по муниципальным образованиям. Набор данных помимо названия субъекта Российской Федерации содержит географические координаты, название населённого пункта и территориальной единицы (района, муниципалитета) внутри субъекта. Это позволяет выполнять более детализированные агрегации и показывать их на карте, например спуститься от субъектов Российской Федерации на уровень ниже и перейти к муниципалитетам. На рисунке 4 показано количество субъектов МСП в муниципальных образованиях России в 2021 г, а также преимущественная специализация муниципалитетов исходя из того, каких фирм в них больше всего.

Отслеживание отдельных фирм. Набор данных позволяет анализировать изменения у конкретных организаций и предпринимателей, например «переезд» (изменение адреса регистрации) из одного региона в другой.

Экономическая, социальная и политическая география
Сёмин П.О.

На рис. 5 показано изменение количества субъектов МСП из пробного набора, зарегистрированных в субъектах Российской Федерации, за весь проанализированный период.

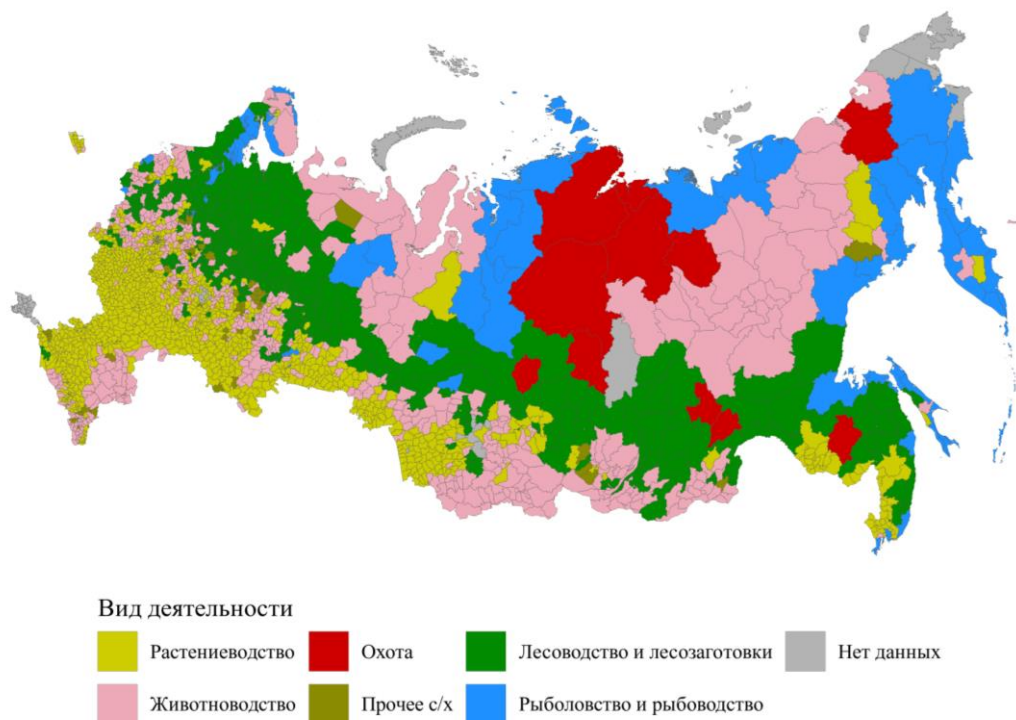


Рис. 4. Число субъектов МСП по муниципальным образованиям и преобладающий вид деятельности. Цвет соответствует тому виду деятельности, который указан в качестве основного у наибольшего числа субъектов МСП в данном муниципалитете (см. легенду). Границы муниципалитетов показаны в соответствии с GADM level2

Fig. 4. The count of SMEs in Russian municipal territories with the dominant activity type, 2021. The most popular activity in the municipality is indicated by color (see legend). GADM municipal borders (level2) are used

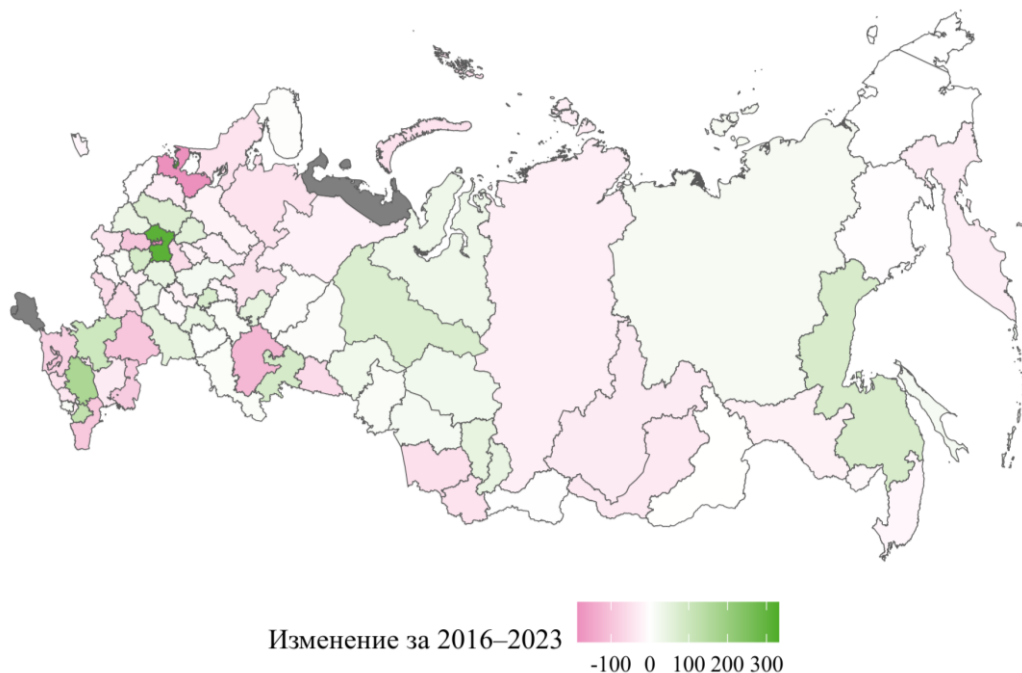


Рис. 5. Распределение количества субъектов МСП из демонстрационного набора, прибывших или покинувших территорию региона (субъекта Российской Федерации) за 2016–2023 гг.

Fig. 5. Regional distribution of SMEs which migrated to or from the region during 2016–2023

Обсуждение. Предложенный алгоритм имеет техническую ценность: он позволяет обработать большой массив сложноструктурированных данных и сделать на его основе несколько сравнительно компактных таблиц в формате, который традиционно используется исследователями. Кроме того, технология нормализует адреса, содержащиеся в исходных данных, и обогащает их географическими координатами и кодами ОКТМО, тем самым делая их пригодными для обработки в геоинформационных системах (ГИС) и пространственного анализа в целом, а также упрощая объединение с муниципальной статистикой. Для выполнения алгоритма, несмотря на использование Apache Spark, достаточно персонального компьютера средней конфигурации – применять облачные вычислительные ресурсы не обязательно. Таким образом, для конечного пользователя технология представляет собой способ работать с big data, обойдя при этом сопутствующие сложности и повышенные требования к ресурсам и навыкам и с помощью нескольких оптимизаций превратив её в работу с обычными «малыми» данными. Благодаря открытому исходному коду и использованию свободного программного обеспечения алгоритм может пригодиться не только пользователям, но и разработчикам как база для создания собственных продуктов.

Набор данных, формируемый с помощью описанной технологии, является информацией о генеральной совокупности субъектов МСП в России. Его можно использовать для сплошного исследования малого и среднего предпринимательства в стране. В некотором роде это квази-источник первичных статистических сведений, которые, по общему правилу, относятся к информации ограниченного доступа, но в данном случае являются открытыми в силу прямого указания в Налоговом кодексе.

Технология де-факто позволяет получить полный перечень индивидуальных предпринимателей с именами и адресами (с точностью до населённого пункта). ФНС России не публикует сведения о доходах, расходах и числе работников применительно к индивидуальным предпринимателям, из-за чего исследовательская ценность этих данных меньше, чем информации об организациях. Также следует отметить, что анализ сведений об индивидуальных предпринимателях может рассматриваться как обработка персональных данных физических лиц, что создаёт правовые риски, особенно актуальные в свете регулярного ужесточения юридической ответственности за нарушения в этой сфере.

Вопрос о том, можно ли использовать технологию и создаваемый с её помощью набор данных для изучения не только малого и среднего бизнеса, но и экономики в целом, является сложным. Как отмечалось в обзоре литературы, потребность в данных, детализированных до уровня отдельных фирм, существует, но получить такую информацию сложно или дорого. Может быть, сведений о субъектах МСП будет достаточно, чтобы делать правдоподобные выводы об экономике в целом? На первый взгляд, это разумное предположение, ведь малые и средние предприятия дают 43 % ВВП, обеспечивают занятостью 38 % работников и составляют почти 98 % в общем числе фирм в России [6]. Столь большое пересечение выборки (реестра МСП) и генеральной совокупности (всех российских юридических лиц и индивидуальных предпринимателей) позволяет предположить, что они во многих случаях взаимозаменяемы, но не гарантирует, что в некоторых ситуациях (отдельные субъекты Российской Федерации или виды деятельности) не будут наблюдаться существенные расхождения. Следовательно, стоит всё же оценить репрезентативность реестра МСП количественно, в том числе с учётом субъекта Российской Федерации или вида деятельности, а не полагаться исключительно на 98-процентное совпадение с множеством всех субъектов экономической деятельности.

Очевидно, что реестр МСП не является случайной выборкой из генеральной совокупности. Причина этого проста: он пополняется не случайно, а исходя из ограничений по финансовым показателям и числу сотрудников. Следовательно, если оценивать репрезентативность реестра МСП по процедуре его формирования, то есть рассуждать теоретически, то он нерепрезентативен: позволяет делать выводы о небольших фирмах, но не распространять их на все компании. Возможен, однако, и альтернативный, эмпирический путь: чтобы понять, является ли реестр МСП достаточно хорошим «прокси» для генеральной совокупности фирм, надо сравнить распределение фирм в нём с распределением в генеральной совокупности. Обычно при анализе репрезентативности сделать такое сравнение невозможно, поскольку характеристики генеральной совокупности неизвестны. В случае с фирмами в России ситуация другая: статистический регистр Росстата представляет собой относительно полный источник сведений о генеральной совокупности.

Для оценки репрезентативности реестра МСП можно использовать корреляционный анализ. Единая межведомственная информационно-статистическая система (ЕМИСС) содержит данные о числе организаций в разрезе субъектов России и групп ОКВЭД за 2019–2022 гг. (<https://fedstat.ru/indicator/58109>). Такая же статистика рассчитывается на базе реестра МСП. В результате для каждого года, каждого региона и каждой группы ОКВЭД известны два числа: общее количество организаций и количество организаций в реестре МСП. Далее вычисляются коэффициенты корреляции по годам, регионам и группам. Используется ранговый коэффициент корреляции Спирмена, так как он более устойчив к выбросам, чем коэффициент корреляции Пирсона.

Результаты анализа показывают высокую согласованность между числом организаций в целом и числом организаций в реестре МСП. Общий коэффициент корреляции в течение рассмотренных четырёх лет составляет чуть больше 0,8 и мало изменяется (табл. 1). Коэффициенты корреляции по регионам различаются довольно сильно (рис. 9), но в основном составляют от 0,7 до 0,9. При анализе отдельных групп ОКВЭД коэффициент корреляции в большинстве случаев не меньше 0,9, а иногда приближается к единице (табл. 2). Лишь три группы

Экономическая, социальная и политическая география
Сёмин П.О.

характеризуются сравнительно небольшой корреляцией: госуправление (О), образование (Р) и экстерриториальная деятельность (U). Это объяснимо, потому что в госуправлении и образовании подавляющее большинство организаций являются государственными или муниципальными (органы власти, подведомственные им учреждения, школы, детские сады, вузы), а количество организаций, ведущих экстерриториальную деятельность, слишком мало, и даже небольшие абсолютные расхождения в числах резко меняют общую картину. Если повторить анализ по годам и регионам, исключив три наиболее «проблемные» группы, то общие коэффициенты корреляции по годам повысятся до 0,94–0,95, а по регионам сместятся в сторону 0,9–1,0. (табл. 1, рис. 9).

Таблица 1

Корреляция между числом организаций в реестре МСП и общим числом организаций в каждом регионе (субъекте Российской Федерации) в каждой группе деятельности по годам
Correlation between the count of SMEs and the total count of commercial organizations by year

Год	$\rho_{Spearman}$, все группы ОКВЭД	$\rho_{Spearman}$, без О, Р, U
2019	0,83	0,95
2020	0,82	0,95
2021	0,81	0,94
2022	0,81	0,94

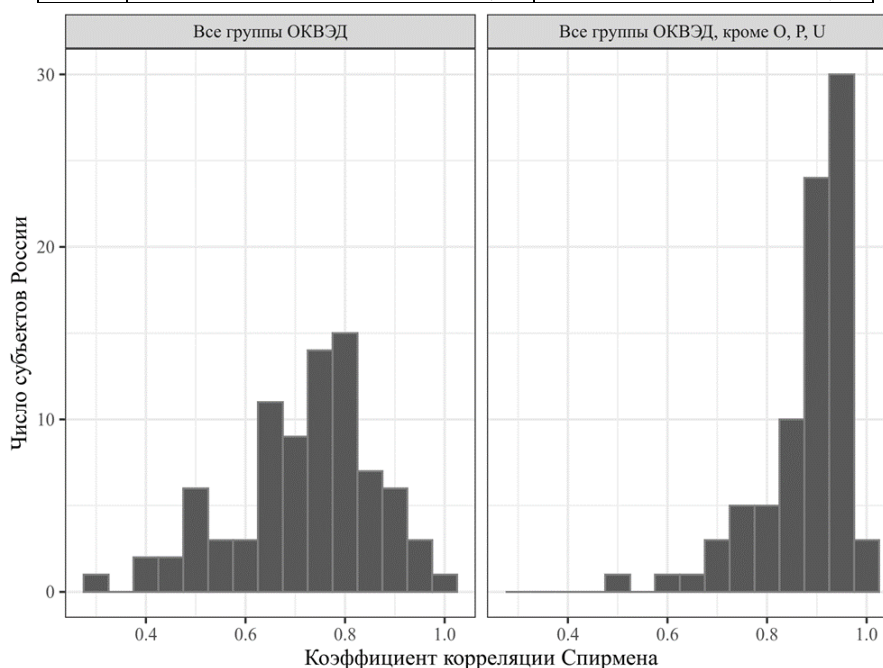


Рис. 6. Распределение регионов (субъектов Российской Федерации) по корреляции между числом организаций в реестре МСП и общим числом организаций в каждом году в каждой группе деятельности

Fig. 6. Distribution of regions by the correlation coefficient between the count of SMEs and the total count of commercial organizations

Таблица 2

Корреляция между числом организаций в реестре МСП и общим числом организаций в каждом регионе (субъекте Российской Федерации) и каждом году по группам ОКВЭД
Correlation between the count of SMEs and the total count of commercial organizations by OKVED activity group

Код группы	Наименование группы	$\rho_{Spearman}$
U	Деятельность экстерриториальных организаций и органов	0,38
О	Государственное управление и обеспечение военной безопасности; социальное обеспечение	0,68
Р	Образование	0,78
Р	Деятельность в области культуры, спорта, организации досуга и развлечений	0,88
Т	Деятельность домашних хозяйств как работодателей; недифференцированная деятельность частных домашних хозяйств по производству товаров и оказанию услуг для собственного потребления	0,89
А	Сельское, лесное хозяйство, охота, рыболовство и рыбоводство	0,93
С	Предоставление прочих видов услуг	0,94
Q	Деятельность в области здравоохранения и социальных услуг	0,96
К	Деятельность финансовая и страховая	0,97

Экономическая, социальная и политическая география
Сёмин П.О.

Окончание табл. 2

Код группы	Наименование группы	ρ_{Spearman}
E	Водоснабжение; водоотведение, организация сбора и утилизации отходов, деятельность по ликвидации загрязнений	0,97
D	Обеспечение электрической энергией, газом и паром; кондиционирование воздуха	0,98
L	Деятельность по операциям с недвижимым имуществом	0,98
B	Добыча полезных ископаемых	0,98
J	Деятельность в области информации и связи	0,99
F	Строительство	0,99
N	Деятельность административная и сопутствующие дополнительные услуги	0,99
G	Торговля оптовая и розничная; ремонт автотранспортных средств и мотоциклов	0,99
M	Деятельность профессиональная, научная и техническая	0,99
I	Деятельность гостиниц и предприятий общественного питания	1,00
C	Обрабатывающие производства	1,00
H	Транспортировка и хранение	1,00

Таким образом, в большинстве случаев реестр МСП может рассматриваться как достаточно качественная замена полному списку организаций как минимум в двух аспектах: он корректно отражает географическое распределение количества организаций по субъектам Российской Федерации в подавляющем большинстве групп ОКВЭД и структуру экономики по большинству групп ОКВЭД в значительной части субъектов Российской Федерации. Результаты корреляционного анализа выглядят весьма очевидными с учётом того, что субъекты МСП – это 98 % всех российских фирм, но позволяют выделить частные случаи, в которых реестр МСП слабо отражает генеральную совокупность: госуправление, образование, экстерриториальная деятельность. Репрезентативность по количеству фирм в разрезе субъектов Российской Федерации и групп ОКВЭД не гарантирует репрезентативности на более низком уровне (конкретные коды ОКВЭД, муниципальные образования, населённые пункты) или по другим характеристикам (число работников, прибыль). Тем не менее технология позволяет подготовить такой набор данных, который можно использовать в ряде задач при отсутствии альтернатив и в условиях дополнительных проверок не только при изучении малого и среднего предпринимательства как такового, но и при географическом исследовании экономической активности всех организаций. Кроме того, набор данных можно применять для разведочного анализа, по результатам которого искать полные данные для более качественной проверки гипотез.

Ещё одной проблемой является различие между реальным и формальным положением дел. Юридический адрес может отличаться от места ведения деятельности, а ОКВЭД, указанный в документах, может не соответствовать фактической деятельности. Сложности усиливаются из-за того, что расхождения могут проявляться в разной степени в зависимости от расположения и отрасли экономики: так, в особых экономических зонах много номинально зарегистрированных организаций, а место деятельности интернет-компаний обычно не совпадает с местом нахождения офиса. Вряд ли эту проблему можно решить системно, поэтому технология не проверяет данные о местонахождении и виде деятельности, предполагая, что они достоверны. В некоторых случаях объём данных сам по себе снижает эффект от неизбежных расхождений. Если возникают сомнения, то целесообразно выборочно проверить соответствие между номинальным и фактическим положением дел. Для этого можно использовать открытые картографические сервисы, сайты организаций и здравый смысл.

Заключение

В статье описана технология, позволяющая на основе открытых данных ФНС России создать геопривязанный набор данных о субъектах МСП (организациях и индивидуальных предпринимателях), ведущих деятельность определённого вида. Получаемый набор охватывает период с августа 2016 г. по настоящее время, содержит сведения о регистрационных номерах, наименовании, местонахождении фирмы, её виде и категории, доходах, расходах и числе работников. В пространственном отношении он детализирован до населённых пунктов, во временном – до лет (в части доходов, расходов и числа работников) или месяцев (по остальным характеристикам), в структурном – до кода вида деятельности по ОКВЭД. Местонахождение фирмы указано как нормализованный адрес населённого пункта, его географические координаты и код ОКТМО. Технология реализована на языке программирования Python с использованием свободного программного обеспечения и позволяет обрабатывать большие (несколько терабайт) объёмы исходных данных на компьютере средней конфигурации за разумное время (обычно не более суток).

Технология может использоваться как готовый способ получения данных для экономико-географических исследований или как база для создания производных решений. Поскольку получаемый набор данных представляет собой сплошной геопривязанный перечень субъектов МСП в России за несколько лет, то он может применяться для пространственного и временного анализа малого и среднего предпринимательства. Благодаря относительной репрезентативности реестра МСП по отношению к генеральной совокупности российских фирм в географическом и структурном плане технология может использоваться как источник данных везде, где требуется

Экономическая, социальная и политическая география
Сёмин П.О.

геопривязанная информация об экономической активности с детализацией до конкретных организаций или населённых пунктов. Технология позволяет самостоятельно получить данные, аналогичные тем, которые содержатся в коммерческих системах бизнес-аналитики или в картографических сервисах. Поскольку создаваемый набор данных содержит координаты для каждой записи (элемента), он удобен также для картографической визуализации.

К ограничениям технологии можно отнести неизбежные расхождения между номинальными сведениями, указанными в реестрах, и фактическим положением дел; охват только части фирм, причём не обязательно репрезентативной во всех отношениях; правовые риски при анализе данных об индивидуальных предпринимателях, которые одновременно являются персональными данными. Из-за неполноты адресных баз данных и изменчивости классификатора муниципального деления геопривязка также происходит корректно не во всех случаях. Кроме того, с учётом тенденции к снижению открытости административных данных есть опасность, что исходные наборы данных на сайте ФНС России перестанут быть доступными.

Библиографический список

1. Graham M., Shelton T. Geography and the future of big data, big data and the future of geography // *Dialogues in Human Geography*. 2013. No. 3 (3). P. 255–261.
2. Liu W. [et al.]. The geography of human activity and land use: A big data approach // *Cities*. 2020. (97). С. 102523.
3. Бабкин Р.А. Оценка численности населения муниципальных образований Московского столичного региона по данным операторов сотовой связи // *Вестник Московского университета. Серия 5. География*. 2020. № 4. С. 116–121.
4. Бабкин Р. А. Опыт использования данных операторов сотовой связи в зарубежных экономико-географических исследованиях // *Вестник Санкт-Петербургского университета. Науки о Земле*. 2021. № 3 (66). С. 416–439.
5. Балина Т. А. [и др.]. Статус территории опережающего развития как механизм управления регионом // *Тихоокеанская география*. 2021. № 2. С. 33–42.
6. Барина В.А., Земцов С.П. Международный сравнительный анализ роли малых и средних предприятий в национальной экономике: статистическое исследование // *Вопросы статистики*. 2019. № 26 (6). С. 55–71.
7. Блануца В.И. Проблемы развития социально-экономического районирования в эпоху больших данных // *Региональные исследования*. 2017. № 2 (56). С. 4–11.
8. Виноградова О.Л. Оценка репрезентативности показателей интенсивности земледелия (на примере стран Европы) // *Географический вестник=Geographical bulletin*. 2020. № 2. С. 62–74.
9. Воробьев А.Н. Использование геоданных для оперативного изучения населения // *Вестник ВГУ. Серия: География. Геоэкология*. 2021. № 2. С. 49–54.
10. Гуменюк И.С. К вопросу о динамике экономической активности и ее влиянии на бюджетную устойчивость муниципальных образований Калининградской области // *Вестник БФУ им. И. Канта – Естественные и медицинские науки*. 2022. № 1. С. 44–56.
11. Кожевников С.А. Модернизация экономики малых городов российского Севера на основе активизации межмуниципальных хозяйственных связей // *Север и рынок: формирование экономического порядка*. 2023. № 3/2023 (26). С. 150–164.
12. Коломак Е.А., Шерубнёва А.И. Оценка влияния агломерационных факторов на экономическую активность (микроэкономический анализ) // *Экономика региона*. 2023. № 3 (19). С. 766–781.
13. Краснов А.И., Краснова М.В. Нестатистические методы оценки трансформации системы расселения Псковской области (по данным сети «ВКонтакте» и сервиса «Яндекс») // *Географический вестник=Geographical bulletin*. 2020. № 4. С. 53–63.
14. Кузьминов И.Ф., Лобанова П.А. Использование текст-майнинга в экономико-географическом отраслевом анализе целлюлозно-бумажной промышленности Европейской России // *Региональные исследования*. 2021. № 1. С. 18–33.
15. Лачининский С.С., Логвинов И.А., Васильева В.А. Оценка спрота городских территорий Санкт-Петербурга на основе спутниковых изображений Landsat // *Вестник Санкт-Петербургского университета. Науки о Земле*. 2023. № 3 (68).
16. Лачининский С.С., Сорокин И.С. Пространственная структура и особенности развития поселений Санкт-Петербургской агломерации // *Балтийский регион*. 2021. № 1 (13). С. 48–69.
17. Макушин М.А. [и др.]. Социально-экономическое развитие территорий в зоне влияния БАМ: советские планы и российские реалии // *Географический вестник=Geographical bulletin*. 2023. № 2 (65). С. 12–25.
18. Моисеева Е.Н., Скугаревский Д.А. Рынок юридических услуг в России: что говорит статистика (Серия «Аналитические обзоры по проблемам правоприменения») / Е.Н. Моисеева, Д.А. Скугаревский. ИПП ЕУСПб, 2016.
19. Никоноров С.М., Кривичев А.И., Максимов Ю.И. Управление социально-экономической политикой в моногородах республики Коми // *Экономика устойчивого развития*. 2021. № 4 (48). С. 123–129.
20. Панкратов А.А. Анализ современного состояния российской ИТ-отрасли: ключевые проблемы и тенденции // *ИнтерКарто. ИнтерГИС*. 2023. № 1 (29). С. 201–216.
21. Петров Ю.В. Пространственное сочетание сельской и городской местности на юге Тюменской области: проблемы, возможные решения // *Географическая среда и живые системы*. 2021. № 3. С. 54–75.
22. Ростислав К.В. Экономико-географическое положение как фактор различий в производительности между регионами России // *Региональные исследования*. 2020. № 3. С. 79–91.
23. Ростислав К.В. Влияет ли географическое сосредоточение на прибыльность российских предприятий? // *Региональные исследования*. 2021. № 1 (71). С. 4–17.

Экономическая, социальная и политическая география

Сёмин П.О.

24. Саранча М.А. Методика оценки уровня и масштабов развития малого предпринимательства в Приволжском федеральном округе (на примере деятельности гостиниц и ресторанов) // Вестник ассоциации вузов туризма и сервиса. 2014. (8). С. 27–32.

25. Сафронов С.Г. Трансформация третичной сферы экономики в регионах России в постсоветский период // Известия Российской академии наук. Серия географическая. 2021. № 4 (85). С. 485–499.

26. Федорченко А.В. Количественная оценка и картографирование территориальных сдвигов в отраслях мирового хозяйства // Вестник Московского университета. Серия 5. География. 2017. № 1. С. 13–19.

References

1. Babkin, R. A. (2020) Estimation of the population of municipalities in the Moscow metropolitan region according to the data of mobile operators. *Lomonosov Geography Journal*, (4), 116–121. (In Russian)

2. Babkin, R. A. (2021) The experience of using the mobile phone data in economic geographical researches in foreign. *Vestnik of Saint Petersburg University. Earth Sciences*, 66(3), 416–439. (In Russian) <https://dx.doi.org/10.21638/spbu07.2021.301>

3. Balina, T. A., Mel'nikov, E. R., Nikolaev, R. S., Stolbov, V. A., Chekmeneva, L. I. (2021) The priority development area as a mechanism of regional management. *Pacific Geography*, (2), 33–42. (In Russian) <https://doi.org/10.35735/tig.2021.6.2.003>

4. Barinova, V. A., Zempsov, S. P. (2019) International Comparative Analysis of the Role of Small and Medium-Sized Enterprises in the National Economy: A Statistical Study. *Voprosy Statistiki*, (26(6)), 55–71. (In Russian) <https://voprstat.elpub.ru/jour/article/view/925>

5. Blanutsa, V. I. (2017) Problems of development of socio-economic regionalization in the era of «big data». *Regional'nye issledovaniia*, (2 (56)), 4–11. (In Russian)

6. Fedorchenko, A. V. (2017) Quantitative evaluation and mapping of territorial shifts in the world economy branches. *Vestnik Moskovskogo universiteta. Seriya 5, Geografiya*, (1), 13–19. (In Russian) <https://vestnik5.geogr.msu.ru/jour/article/view/270>

7. Graham, M., Shelton, T. (2013) Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography*, 3(3), 255–261. <https://dx.doi.org/10.1177/2043820613513121>

8. Gumenyuk, I. S. (2022) On the dynamics of economic activity and its impact on the budgetary stability of municipalities of the Kaliningrad region. *IKBFU's Vestnik. Series: Natural and Medical Sciences*, (1), 44–56. (In Russian)

9. Kolomak, E. Anatol'evna., Sherubneva, A. I. (2023) Assessment of the Impact of Agglomeration Factors on the Economic Activity: Microeconomic Analysis. *Ekonomika regiona / Economy of regions*, 19(3), 766–781. (In Russian) <https://dx.doi.org/10.17059/ekon.reg.2023-3-12>

10. Kozhevnikov, S. A. (2023) Upgrading the economies of small towns in the Russian North through enhanced intermunicipal economic cooperation. *Sever i rynek: formirovanie ekonomicheskogo poriyadka [The North and the Market: Forming the Economic Order]*, 26(3/2023), 150–164. (In Russian) <https://dx.doi.org/10.37614/2220-802x.3.2023.81.010>

11. Krasnov, A. I., Krasnova, M. V. (2020) Non-statistical methods for assessing the transformation of the settlement system of the Pskov region (according to the VKontakte network and the Yandex service). *Geographical bulletin*, (4), 53–63. (In Russian) <http://press.psu.ru/index.php/geogr/article/view/4031>

12. Kuzminov, I. F., Lobanova, P. A. (2021) Text mining for economic geographical sectoral analysis of the pulp and paper industry in European Russia. *Regional'nye issledovaniia*, (1), 18–33. (In Russian)

13. Lachininskii, S. S., Logvinov, I. A., Vasileva, V. A. (2023) Assessment of urban sprawl of St. Petersburg urban areas based on Landsat satellite images. *Vestnik of Saint-Petersburg University. Earth Sciences*, 68(3). (In Russian) <https://dx.doi.org/10.21638/spbu07.2023.303>

14. Lachininsky, S. S., Sorokin, I. S. (2021) Spatial structure and development of settlements in the Saint Petersburg agglomeration. *Baltic region*, 13(1), 48–69. (In Russian) <https://dx.doi.org/10.5922/2079-8555-2021-1-3>

15. Liu, W., Wu, W., Thakuriah, P., Wang, J. (2020) The geography of human activity and land use: A big data approach. *Cities*, 97, 102523. <https://dx.doi.org/https://doi.org/10.1016/j.cities.2019.102523>

16. Makushin, M. A., Bobrovskiy, R. O., Demidova, K. V., Danshin, A. I., Goryachko, M. D. (2023) Socio-economic development of the BAM territories: Soviet plans and Russian realities. *Geographical bulletin*, (2(65)), 12–25. (In Russian) <https://dx.doi.org/10.17072/2079-7877-2023-2-12-25>

17. Moiseeva, E. N., Skougarevskiy, D. A. (2016) *Rynek iuridicheskikh uslug v Rossii: chto govorit statistika (Serii «Analiticheskie obzory po problemam pravoprimereneniia»)*. IRL EUSPb Publ.

18. Nikonorov, S. M., Krivichev, A. I., Maksimov, I. I. (2021) Management of socio-economic policy in monotowns of the Komi republic. *Economics of sustainable development*, (4 (48)), 123–129. (In Russian) https://doi.org/10.37124/20799136_2021_4_48_123

19. Pankratov, A. (2023) Analysis of the current state of the Russian IT Industry: key issues and trends. *InterCarto. InterGIS*, 29(1), 201–216. (In Russian) <https://dx.doi.org/10.35595/2414-9179-2023-1-29-201-216>

20. Petrov, Y. V. (2021) Spatial combination of rural and urban areas in the south of the Tyumen region: problems and possible solutions. *Bulletin of the Moscow State Regional University (Geographical Environment and Living Systems)*, (3), 54–75. (In Russian) <https://dx.doi.org/10.18384/2712-7621-2021-3-54-75>

21. Rostislav, K. V. (2021) Does geographical concentration affect the profitability of Russian enterprises? *Regional'nye issledovaniia*, (1 (71)), 4–17. (In Russian)

22. Rostislav, K. V. (2020) Economic-geographical location as a factor in productivity differences between Russia's regions. *Regional'nye issledovaniia*, (3), 79–91. (In Russian)

Экономическая, социальная и политическая география
Сёмин П.О.

23. Safronov, S. G. (2021) Transformation of the Tertiary Sphere of Russia's Regions Economy in the Post-Soviet Period. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Geograficheskaya*, 85(4), 485–499. (In Russian) <https://dx.doi.org/10.31857/s2587556621040087>

24. Sarancha, M. A. (2014) The methodology for estimating the level and scale of small businesses development in Volga Federal District (on the example of hotels and restaurants). *Universities for Tourism and Service Association Bulletin*, 8, 27–32. (In Russian)

25. Vinogradova, O. L. (2020) Agricultural intensity indicators: representativity assessment (a case study of European countries). *Geographical bulletin*, (2), 62–74. (In Russian) <http://press.psu.ru/index.php/geogr/article/view/3271>

26. Vorobyev, A. N. (2021) Use of Geodata for Operational Study of Population. *Vestnik Voronezskogo gosudarstvennogo universiteta. Seria: Geografia. Geoekologia*, (2), 49–54. (In Russian) <https://dx.doi.org/10.17308/geo.2021.2/3447>

Статья поступила в редакцию: 10.04.24, одобрена после рецензирования: 15.11.2024, принята к опубликованию: 12.12.2024.

The article was submitted: 10 April 2024; approved after review: 15 November 2024; accepted for publication: 12 December 2024.

Информация об авторе

Павел Олегович Сёмин

аспирант

Пермский государственный

национальный исследовательский университет,

бэкэнд-разработчик ООО «Агро-Софтвр»;

614068, Россия, г. Пермь, ул. Букирева, 15

Information about the author

Pavel O. Syomin

PhD Student

Perm State University;

Backend Developer at AgroSoftware LLC;

15 Bukireva st., Perm, 614068, Russia

e-mail: ntsp@ya.ru