

УДК 338**БАЗЫ ДАННЫХ МИНИСТЕРСТВ И ВЕДОМСТВ КАК ИСТОЧНИК
ВАЖНОЙ ИНФОРМАЦИИ О ЛОКАЛЬНОЙ СЕЛЬСКОЙ
ЭКОНОМИКЕ****DATABASES OF MINISTRIES AND DEPARTMENTS AS A SOURCE OF
IMPORTANT INFORMATION ON THE LOCAL RURAL ECONOMY***Сальников С.Г.,¹**Тухина Н.Ю.²*

Рассматриваются базы данных различных министерств, в которых можно найти информацию по локальной сельской экономике (ЛСЭ). Показана важная роль этой информации для изучения ЛСЭ. Рассмотрены также межведомственные источники информации по ЛСЭ. Представлены статистические сведения по указанной информации.

There was information on the local rural economy in databases of different ministries and departments. The important role of this information for study was demonstrated. Interdepartmental sources of information were also considered. Statistics on the information has been provided.

Ключевые слова: *базы данных, локальная сельская экономика, межведомственные источники информации.*

Key words: *date base, local rural economy, interdepartmental sources of information.*

DOI 10.37691/2619-0265-2022-0-2-109-122

¹ **Сальников Сергей Георгиевич**, к. ф.-м. н., ведущий научный сотрудник, руководитель отдела информатизации АПК, Всероссийский институт аграрных проблем и информатики имени А.А. Никонова - филиал Федерального государственного бюджетного научного учреждения «Федеральный научный центр аграрной экономики и социального развития сельских территорий - Всероссийский научно-исследовательский институт экономики сельского хозяйства», г. Москва, e-mail: salns@gmail.com

Salnikov Sergey Georgievich, Candidate of Physical and mathematical Sciences, leading researcher, Head of the Agribusiness Informatization Department, All-Russian Institute of Agrarian Problems and Informatics named after A.A. Nikonov-Branch of the Federal State Budgetary Scientific Institution «Federal Research Center of Agrarian Economy and Social Development of Rural Areas - All-Russian Research Institute of Agricultural Economics», Moscow, e-mail: salns@gmail.com

² **Тухина Наталья Юрьевна**, научный сотрудник отдела информатизации АПК, Всероссийский институт аграрных проблем и информатики имени А.А. Никонова-филиал Федерального государственного бюджетного научного учреждения «Федеральный научный центр аграрной экономики и социального развития сельских территорий-Всероссийский научно-исследовательский институт экономики сельского хозяйства», г. Москва, e-mail: tukhina@viapi.ru

Tukhina Natalia Yurievna, researcher of Agribusiness Informatization Department, All-Russian Institute of Agrarian Problems and Informatics named after A.A. Nikonov - Branch of the Federal State Budgetary Scientific Institution «Federal Research Center of Agrarian Economy and Social Development of Rural Areas - All-Russian Research Institute of Agricultural Economics», Moscow, e-mail: tukhina@viapi.ru

Базы данных министерств и ведомств – один из ключевых источников информации по ЛСЭ. Изучение этих источников поможет в получении важной информации по ЛСЭ. Изучение представленности информации по ЛСЭ, её полноты, периодичности, форматов представления – одна из основных целей представленного ниже анализа данных. Будут рассмотрены не только ведомственные источники данных, но и межведомственные источники, прежде всего ЕМИСС (Единая межведомственная информационно-статистическая система). Также краткая информация будет представлена по БД ПМО (базе данных показателей муниципальных образований).

База данных показателей муниципальных образований (БД ПМО). Один из наиболее важных и полных источников по муниципальным образованиям, в том числе в сельской местности. Эта БД содержит т.н. муниципальную статистику за период с 2006 по 2020 гг.

Данная БД обладает рядом существенных преимуществ, в первую очередь большим объёмом представленных показателей, глубина разреза – вплоть до населённых пунктов, простотой представления, большой периодичностью и скоростью вывода данных и т.п.

Наряду с этим БД ПМО обладает и рядом существенных недостатков [1]:

- Хотя доступ к БД ПМО открытый, но скачивание данных из этой БД возможно только в строго ограниченном (и единственном) формате. Скачивание данных также зачастую заканчивается сообщением об ошибке из-за каких-то (в целом нам неизвестных) ограничений по объёму скачиваемых данных
- Выходные данные в виде таблиц зачастую имеют «лакуны», то есть незаполненные ячейки таких таблиц.

Изучению представленности «лакун» в БД ПМО и способов преодоления этого недостатка БД ПМО и посвящена большая часть статьи [1]. Показано, что эта БД достаточно полна для 2011-2019 гг. Показано также, что при использовании достаточно простых алгоритмов заполнения пропусков в данных (т.н. импутация) они в БД ПМО становятся полными и за 2020 год.

БД ПМО обладает также ещё одним существенным недостатком – не очень удобным (для последующего использования) форматов представления выходных данных. Однако, использование вполне официальной копии этой БД [2] делает данную проблему решаемой.

Базы данных Всероссийских сельхозпереписей 2006 г., 2016 г. и сельскохозяйственной микропереписи 2021 г.

Данные по сельхозпереписи - несомненно, один из самых больших, интересных и полезных источников по сельской локальной экономике. Особенно это касается данных по ЛПХ – личным подсобным хозяйствам, которые сосредоточены территориально в основном в сельской местности.

Большинство собираемой в ходе Переписей информации носит экономический характер. Для примера приведём состав переписного листа для ЛПХ в 2021 году:

1. Общая характеристика
2. Земельные ресурсы
3. Посевные площади сельхозкультур, теплиц и многолетников
4. Использование земельных ресурсов
5. поголовье сельскохозяйственных животных

Как видно, большая часть информации относится к посевам и скоту, хотя есть и другая информация.

Сведения, собранные в ходе переписей публикуются на специализированных сайтах и на сайте Росстата в отдельном разделе сайта [3].

Здесь можно найти сведения по всем трём проведённым в последнее время Переписям – 2006, 2016 и 2021 гг.

Помимо самих итогов переписей приводятся нормативно-правовые документы, протоколы заседаний, образцы форм отчётности и указания по их заполнению, материалы по информационно-разъяснительной работе и др.

К сожалению, итоги переписей пока что можно найти только в агрегированном виде (до уровня регионов) и только в PDF-формате.

Хотя ничего не мешает нам поискать сведения по сельхозпереписям на региональных сайтах Росстата. Здесь нас, правда, ожидают другие сюрпризы представления:

- Представление такой информации не только в PDF-формате, но и виде электронных таблиц и/или текстовых документов;
- Отсутствие единой структуры (для регионов) представления таких данных;
- Существенные пробелы в данных, как на уровне регионов, так и вообще по регионам (вплоть до полного отсутствия сведений по сельхозпереписи для конкретных отдельных регионов и/или годов переписи).

Всё это существенно затрудняет исследования на уровне регионов и тем более муниципальных районов и городских/сельских поселений.

Так, например, для сельхозпереписи 2006 года какие-то сведения (иногда почти полные, но зачастую и весьма фрагментарные) по этой переписи удалось обнаружить только для 55 (из 85) регионов страны.

Намного лучше ситуация с сельхозпереписью 2016 года, где региональные сведения по этой переписи удалось обнаружить на 83 (из 85) сайтах регионов. Однако и здесь сведения неоднородны, зачастую

отсутствуют конкретные тома переписи и/или таблицы. Всё же в целом ситуация с региональными данными по переписи 2016 года намного лучше таковой для 2006 года и близка к идеальной.

Итоги Сельхозпереписи-2006 года выложены на сайте Росстата в виде 9 томов (некоторые тома состоят из нескольких книг) в PDF-формате. Даже таблицы и картограммы выложены в том же формате. Это существенно затрудняет использование данных материалов для исследователей, так как перевод файлов PDF-формата в обычные текстовые файлы или таблицы далеко не так прост и однозначен, как хотелось бы.

Итого Сельхозпереписи-2016 также выложены в виде томов (8 штук) PDF-формата. Для них остаются те же проблемы, что указаны нами выше для Сельхозпереписи-2006. Заметим, что эти Итоги на сайте Росстата также можно найти только в PDF-формате.

Последняя Микросельхозперепись-2021 года пока также представлена только в виде предварительных итогов в PDF-формате (всего, по сельхозорганизациям и по фермерам). Но есть большая вероятность того, что результаты данной Переписи впервые будут представлены в виде файлов конечного доступа (на уровне единиц переписи) – конечно же, в обезличенном виде. Ориентировочная дата представления таких данных – осень 2022 года.

Базы Всероссийских переписей населения 2002, 2010 г. и 2020 г.

Несомненно, эти информационные источники – одни из важных для локальной сельской экономики. На сайте Росстата по переписи населения представлена следующая информация:

1. Всероссийская перепись населения 2020 года;
2. Пробная перепись населения 2018 года;

3. Социально-демографическое обследование (микроперепись населения) 2015 года;
4. Перепись населения в Крымском федеральном округе 2014 года;
5. Всероссийская перепись населения 2010 года;
6. Всероссийская перепись населения 2002 года.

Нас, конечно, прежде всего, интересуют результаты Всероссийских переписей 2002, 2010 и 2020 гг.

Данные Переписи-2002 представлены на отдельном сайте, сгруппированные в тома (14 штук).

Каждый том, однако, не выводится целиком, а может быть получен в виде отдельных таблиц (в XLS, HTML или отдельном архивированном ZIP-файле, в котором заархивирован XLS-файл). Это немного усложняет доступ к данным Переписи, но в целом не представляет особой проблемы.

Другое дело, что каких-то машиночитаемых форматов (XML, JSON и т.п.) на сайте не представлено.

Информация по переписи Всероссийской переписи населения 2010 года представлена в отдельном разделе сайта Росстата. Информация также дана в виде томов (11 томов). Кроме того информацию по Переписи-2010 можно также получить в картографическом представлении. Отдельной колонкой идёт информация по Москве и Московской области (в границах на 1 июля 2012 года). Здесь же можно найти методологические пояснения и образцы переписных листов.

Всероссийская перепись 2020 года также представлена в отдельном разделе сайта Росстата [4].

Здесь, помимо предварительных итогов Переписи, можно найти нормативно-правовые документы по Переписи, протоколы заседания Комиссии по проведению Переписи, информационно-разъяснительные материалы, образцы переписных листов и пр.

Предварительные итоги Переписи-2020 представляют собой 2 страницы текста в DOC-формате с разъяснениями и одну (!) небольшую электронную таблицу с общей численностью населения по регионам в сопоставлении с такими же данными Переписи 2010 года. Так что информации по Переписи-2020 ещё представлено очень мало, и следует ожидать появления последующей информации.

Базы данных Министерства здравоохранения

Информация по сельской местности обнаружена нами на сайте Минздрава за 2014-2018 гг. в разделе [5].

Более ранней и более поздней информации нами не обнаружено.

Основная статистическая информация Министерства представлена на сайте Центрального научно-исследовательского института организации и информатизации здравоохранения» Министерства здравоохранения Российской Федерации [6].

К сожалению, доступ к статистической информации на этом сайте возможен только при авторизации в качестве сотрудника какого-либо медицинского информационно-аналитического центра (МИАЦ), так что мы ограничились только теми аналитическими материалами, которые свободно доступны на сайте Минздрава (по ссылке в начале этого подраздела).

За 2018 год найдено 8 бюллетеней с информацией по здоровью, три файла из них содержат информацию по сельской местности; за 2017 год найдено 20 файлов с информацией по здоровью населения, 5 файлов из них содержат информацию по сельской местности; за 2016 год найден 21 бюллетень, из них 6 бюллетеней с данными по сельской местности; за 2015 год представлено 22 бюллетеня, среди которых 6 бюллетеней содержали информацию о сельской местности; За 2014 год представлен 21 бюллетень, среди которых 5 бюллетеней содержали информацию о сельской местности.

Таким образом, имеем следующий свод по имеющейся на сайте Минздрава информации по сельской местности (таб. 1):

Таблица 1

Информация по сельской местности (по данным Минздрава)

Год	Количество таблиц с информацией по сельской местности
2012	17
2014	47
2015	34
2016	30
2017	30
2018	5
ИТОГО	146 таблиц с информацией по ЛСЭ

Источник: расчёты автора

Все представленные по сельской местности данные Минздрава имеют формат Microsoft Office Word (docx-формат), так что их преобразование в формат, подходящий для загрузки в типовые базы данных, потребует значительных усилий.

Базы данных Министерства культуры

Статистическую информацию по Министерству культуры можно найти на сайте АИС «Статистика» ГИВЦ Минкультуры России по адресу [7].

Здесь представлены общие сведения (в основном посещения тех или иных организаций культуры), сведения о цирках, зоопарках, театрах, музеях, библиотеках, парки, клубы, школы искусств и др.

Информация на сайте разделена на подразделы: общие сведения, своды, справочники. К сожалению, такое подразделение имеет место только с 2016 года. До 2016 года информация идёт единым списком и понять, что же представлено в файлах можно только косвенно, сопоставляя их с данными после 2016 года.

Сводная таблица по статистической информации Минкультуры представлена ниже (таб. 2):

Таблица 2

Представленность статистической информации Минкультуры за 2012-2021 гг.

Вид данных	Количество таблиц с сельской информацией (по годам)									
	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Библиотеки (сводь)	-	-	-	-	1	1	1	1	2	4
ДШИ (сводь)	-	-	-	-	5	-	-	-	3	3
Зоопарки (сводь)	-	-	-	-	-	-	-	-	-	-
Клубы (сводь)	-	-	-	-	1	1	2	2	2	2
Концертные организации	-	-	-	-	1	-	2	-	-	-
Музеи (сводь)	-	-	-	-	1	1	2	-	2	2
Театры (сводь)	-	-	-	-	1	1	-	2	1	-
Цирки (сводь)	-	-	-	-	1	-	-	-	-	-
Библиотеки (справочник)	1	1	1	1	1	5	5	5	7	-
Клубы и парки (справочник)	6	6	6	6	6	6	6	6	10	-
Учебные заведения и ДШИ (справочник)	4	4	-	4	-	4	3	6	6	6
ИТОГО по году	11	11	7	11	18	19	21	22	33	17
ИТОГО по всем годам	170 электронных таблиц или файлов PDF-формата									

Источник: расчёты автора

Статистические сведения представлены за 2012-2021 гг. в виде электронных таблиц или PDF-файлов, так что они, вообще говоря, потребуют значительных усилий для преобразования их в записи типовой баз данных.

Базы данных Минпросвещения и Миннауки

Информация по этим Министерствам сосредоточена как на сайтах этих Министерств, так и на сайте Росстата в разделе [8].

Здесь по сельской местности представлены, прежде всего, общая информация:

- охват детей дошкольным образованием
- охват по возрасту детей
- Сборник «Образование в России» за 2021 год (к сожалению, за другие годы этот сборник обнаружить не удалось)

Также есть сводная информация: (уже собранная по регионам) по сельской местности в различных формах:

1-ДОП – дошкольное образование (36 таблиц по типам поселения в 2019 году, 39 таблиц – в 2018 году, 42 таблицы в 2017 году и 43 таблиц в 2016 году)

1-ОЛ – Сведения об организации отдыха детей и их оздоровления (13 таблиц по типам поселения в 2021 году, 12 таблиц в 2020 году, 10 таблиц в 2019 году и 9 таблиц в 2018 году)

85-К - Сведения о деятельности организации, осуществляющей образовательную деятельность по образовательным программам дошкольного образования, присмотр и уход за детьми (44 таблицы с информацией по сельской местности за 2020 год, 31 таблица – за 2019 год, 36 таблиц – за 2018 год, 30 таблиц за 2017 год и 33 таблицы за 2016 год)

Также значительное количество информации по сельской местности можно найти непосредственно на сайте Минпросвещения (на сайте Миннауки и образования данных по сельской местности нет):

таблицы с формой 1-ДО «Сведения об учреждениях дополнительного образования для детей» (за 2013-2021 гг.)

таблицы с формой 1-НД «Сведения о численности несовершеннолетних в возрасте 7-17 лет, не обучающихся в организациях, осуществляющих образовательную деятельность» (за 2014-2022 гг.)

таблицы с формой Д-13 «Сведения об учреждениях для детей-сирот и детей, оставшихся без попечения родителей» (за 2016-2021 гг.)

таблицы с формой ОО-1 «Сведения об организации, осуществляющей образовательную деятельность по образовательным программам начального, общего, основного общего, среднего общего образования» (за 2017-2021 гг.)

таблицы с формой ОО-2 «Сведения о материально-технической и информационной базе, финансово-экономической деятельности общеобразовательной организации» (за 2018-2021 гг.)

таблицы с формой ПО «Сведения о деятельности организации, осуществляющей образовательную деятельность по основным программам профессионального обучения» (за 2018-2021 гг.)

К сожалению, эта информация на сайтах представлена в виде в точности тех же файлов (по формату и по структуре), которые используются при сборе данных, так что преобразование этих файлов в форматы, пригодные для их последующей загрузки в базы данных представляется достаточно сложной и кропотливой работой.

Кроме того, на сайте Минпросвещения есть раздел с открытыми данными. На нём представлено 93 набора данных, которые можно скачать в 5 различных форматах (в том числе и CSV). К сожалению, большая часть этих данных к информации по сельской местности не относится.

Межведомственные базы данных

Здесь мы остановимся только на одном таком информационном источнике по сельской местности – т.н. ЕМИСС (Единая межведомственная информационно-статистическая система) [9].

Эта система содержит в себе статистическую информацию различных агентств и ведомств России в единой форме. Это существенно упрощает доступ в данной информации. Кроме того возможен вывод информации в едином (специализированном, статистическом) формате – SDMX, что также даёт данному сайту явное преимущество перед конкурентами, так как такой подход, например, позволяет без значительных дополнительных затрат объединять в единые информационные массивы статистическую информацию разных ведомств.

Вся информация в ЕМИСС представлена в виде т.н. **«гиперкубов»**. То есть к каждому конкретному показателю привязано конечное (но фиксированное для данного набора данных) количество т.н. осей гиперкуба. В качестве таковых обычно выступают т.н. кодификаторы, например ОКАТО, ОКТМО, справочник товаров и т.п. Задание значений осей гиперкуба однозначно определяет показатель (точку) внутри этого куба.

Среди недостатков данного представления отметим, что информацию в виде SDMX-файлов (в отличие от XLS-файлов) нет возможности сразу вывести целиком на выгрузку, что существенно усложняет возможности работы с этим форматом (статистических) данных.

Есть при таком подходе к работе с выходными файлами и одна явная и очевидная ошибка, о которой не надо забывать. А именно: размер XLS-файла ограничен размером чуть более чем в 64 тыс. строк. Так что, если мы выводим файл в XLS-формате и задали вариант вывода всего файла, то вовсе не факт, что мы получим действительно весь файл целиком, а не его часть!

Нами проведено значительное предварительное исследование того, какие именно информационные источники с этого сайта можно использовать для создания БД с информацией по сельской местности. В результате информационные источники собраны в один (единый) массив информации по сельской местности в соответствии с одним из следующих принципов:

слово **«сельское»** (или его близкие по форме слова) встречается в названии информационного массива (таких массивов-таблиц набралось 70)

справочник **«тип местности»** встречается среди справочников соответствующих информационных массивов (гиперкубов). Таких массивов набралось 170.

В ходе последующего сбора информации значительная часть массивов была отброшена по тем или иным причинам (как правило, это существенная

неполнота данных в конкретном массиве) и всего в окончательном наборе SDMX-файлов было представлено 160 файлов.

Таким образом, ЕМИСС представляет значительный и очень важный массив информации для ЛСЭ.

Заключение

Выше нами представлен только сравнительно небольшой (по объёму) анализ информации в БД различных министерств и ведомств, которые имеют отношение к ЛСЭ. Как видно, информация очень неоднородна, как по составу и временному (и другим) срезам, так и по форматам, периодичности представления и т.п.

Значительный интерес представляет, прежде всего, БД ПМО, где – единственно среди всех рассмотренных БД – представлен срез информации до уровня муниципальных образований и даже населённых пунктов.

Также очень важная информация представлена в ЕМИСС. Ценность её не только, как правило, в разнообразных срезах информации, но и, опять-таки, как правило, в значительной временной длине представленной информации и особом формате (SDMX) представления, что даёт возможность достаточно унифицировано извлекать большие массивы данных, в том числе и по ЛСЭ.

Остальные рассмотренные базы данных имеют ярко выраженную ведомственную специфику. Ситуация осложняется, как видно из анализа выше, только неполнотой таких данных, не очень большой их периодичностью и большим разбросом в форматах представления этих данных. Это не мешает, однако, проанализированным базам данных оставаться важными источниками информации по ЛСЭ.

Литература

1. Сальников, С. Г. База данных показателей муниципальных образований (БД ПМО) как источник полной информации о локальной сельской экономике / С.Г. Сальников, Н.Ю. Тухина // Вестник Московского гуманитарно-экономического института. – 2022. – № 1. – С. 267-280. – DOI 10.37691/2311-5351-2022-0-1-267-280.
2. ИНИД - <http://data-in.ru/data-catalog/datasets/115/>
3. <https://rosstat.gov.ru/folder/75792>
4. https://rosstat.gov.ru/vpn_popul
5. <https://minzdrav.gov.ru/ministry/61/22/stranitsa-979/statisticheskie-i-informatsionnye-materialy/statisticheskie-materialy>
6. <https://mednet.ru/>
7. <https://stat.mkrf.ru/indicators/>
8. <https://rosstat.gov.ru/statistics/education>
9. <https://www.fedstat.ru/>