

«Демографические исследования и экономика труда в цифровой экономике»
Научный семинар по исследованиям цифровой экономики 6 июня 2018 г.

К вопросу о цифровизации, новых технологиях и новых данных в демографии (на примере статистики миграции)



О. ЧУДИНОВСКИХ

ЛАБОРАТОРИЯ ЭКОНОМИКИ НАРОДОНАСЕЛЕНИЯ И ДЕМОГРАФИИ

ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ МГУ ИМЕНИ М.В.ЛОМОНОСОВА

Что мы подразумеваем , говоря о цифровизации и инновационных подходах в демографии в целом и миграциологии в частности? В первую очередь речь идет о прорыве в области данных (и статистики):

Виды данных (новые)

Т.к. существующие не удовлетворяют (по разным причинам) , новые ресурсы интересно протестировать на предмет использования,

Технологии сбора, хранения, обработки и производства статистики

Производительность и оперативность прежних - не отвечают потребностям, они не позволяют интегрировать разные данные, морально устарели

Форма доступа пользователей к информации и распространения данных

- * Бумажные носители и данные в не машиночитаемых форматах морально устарели ,
- * Деперсонифицированная информация , не являющаяся предметом защиты, а тем более- агрегированная - должна быть доступна обществу

Все это не только объективные реалии и потребность времени, но и модный тренд

Основные производители статистики миграции и смежных явлений в России



Листок статистического учета мигранта – основа государственной статистики миграционных потоков. Применяется (с модификациями по формату и перечню переменных) с начала 1930-х годов .

Листки поступают из органов МВД в тер. органы статистики. Не машиночитаемые . Вручную - подготовка к вводу и ввод в органах статистики. Огромные затраты рутинного ручного труда в контексте разговоров о цифровизации всего.

Приложение № 12
Форма № 12П

«П» ЛИСТОК СТАТИСТИЧЕСКОГО УЧЕТА ПРИБЫТИЯ
(к документам о регистрации по новому месту жительства или по месту пребывания)

Записи в листке статистического учета прибытия подлежат использованию только для получения сводных данных о численности и составе мигрантов и относятся к категории конфиденциальной информации

Регистрация: по месту нового жительства в том числе в связи с изменением гражданства по месту пребывания на срок

с «___» _____ (число) _____ (месяц) _____ (год) _____
по «___» _____ (число) _____ (месяц) _____ (год) _____

1. Фамилия	
2. Имя	
3. Отчество	
4. Дата рождения	«___» _____ (число) _____ (год) _____
5. Место рождения	государство _____ республика, край, область, округ _____ район, городской район (округ) _____ город, поселок городского типа _____ сельский населенный пункт _____
6. Пол (подчеркнуть): мужской — 1; женский — 2	
7. Гражданство (указать государство)	
если имеет двойное гражданство, указать государство	
если изменено гражданство, указать предыдущее гражданство (государство)	
8. Новое место жительства	республика, край, область, округ _____ район, городской район (округ) _____ город, поселок городского типа _____ сельский населенный пункт _____
9. Последнее место жительства	государство _____ республика, край, область, округ _____ район, городской район (округ) _____ город, поселок городского типа _____ сельский населенный пункт _____
10. Проживал по последнему месту жительства с _____ года	

Форма № 12П
(продолжение)

11. Основное обстоятельство, вызвавшее необходимость переселения (подчеркнуть)	в связи с учебой — 10; в связи с работой — 20; возвращение к прежнему месту жительства — 30; из-за обострения межнациональных отношений — 40; из-за обострения криминогенной обстановки — 50; экологическое неблагополучие — 60; несоответствие природно-климатическим условиям — 70; причины личного, семейного характера — 80, в т.ч. в связи с переменной места работы супруга(и) — 81, в связи с вступлением в брак — 82, к детям — 83, к родителям — 84; иная причина (указать) _____ — 90, в т.ч. приобретение жилья (покупка, наследование т.п.)
--	---

Давно планировался переход к электронным формам передачи индивидуальных данных о мигрантах на основе ППО «Территория» из органов ФМС (ныне – МВД) в Росстат. 5 апреля 2016 г. - передача функций ФМС России МВД, создание ГУВМ МВД. Приостановка (почти на два года) развития имевшихся систем, закрытие доступа, передача всех автоматизированных систем учета в ДИТСиЗИ МВД, смена команды (связанной с работой этих систем) в центральном аппарате, сокращение и замещение специалистов в регионах. Только недавно удалось вернуться к вопросу предоставления первичной информации в электронном виде, только по иностранным гражданам и по очень ограниченному кругу переменных.

Результаты обращений (официальное письмо) в ФОИВы за статистическими данными в 2017-2018 гг., которых нет в открытом доступе (и которые не являются предметом защиты)

ГУВМ МВД России –

Стандартный запрос на регламентные таблицы 1- и 2-РД. Первый раз – отказ. Второй раз - была предоставлена выборка показателей (без АППГ). С третьей попытки (письмо за подписью проректора МГУ) получены полные таблицы.

Запрос на выгрузку из Центрального банка данных по учету иностранных граждан (ЦБДУИГ) распределений по полу и возрасту лиц получивших виды на жительство и разрешения на временное проживание, патенты и разрешения на работу, гражданство.

Срок ожидания превысил два месяца. Получены в формате Excel .

МИД России - запрос статистики приема в гражданство и консульского учета. Срок ожидания - более двух месяцев. Получено не в машиночитаемом формате (на бумажном носителе в виде фотографии/скана) .

Сведения о количестве граждан Российской Федерации, состоявших на консульском учете за рубежом в 2012-2016 гг.

Страна	2012			2013			2014			2015			2016			Всего
	Временное проживание	Постоянное проживание	Визы													
Австрия	399	37	431	843	48	911	1250	117	1367	1056	265	2301	2306	2607	2572	
Азербайджан	170	8792	8762	195	3660	6660	261	6180	6441	267	6312	6562	291	6172	6463	
Австралия	636	1808	2444	652	1924	2699	706	2126	2822	709	2262	3091	686	2541	3227	
Албания	3354	6497	12811	3170	9649	13219	3363	9982	11365	3569	9712	13081	3345	10302	13647	
Алжир	138	252	396	176	266	426	168	238	366	156	226	316	138	188	278	
Англия	521	692	1044	553	901	1054	637	511	1148	830	339	1469	763	459	1222	
Аргентина	656	23	479	112	25	527	874	31	465	667	34	701	691	24	736	
Армения	664	562	1286	961	629	1521	1013	629	1635	1017	626	1657	1018	696	1666	
Аustria	478	17817	18295	507	16489	18991	551	19591	20923	640	20119	20679	593	21325	21918	
Афганистан	309	97	666	409	101	510	466	109	576	492	95	567	614	53	249	
Багамские	94	21	115	96	21	126	101	21	122	102	11	163	136	34	176	
Бангладеш	180	34	284	180	80	270	192	71	263	141	71	212	141	88	202	
Барбадос	1766	60272	60665	6466	50706	64276	13821	60366	64076	4324	62947	67271	5630	10181	39711	
Бразилия	6665	2742	9407	8363	3296	11671	9232	3773	13091	9187	4410	13597	9190	4420	14010	
Бразилия	32	131	183	23	149	172	122	41	166	2	13	35	4	35	17	
Бразилия и Германия	155	168	261	158	115	272	109	137	308	72	223	366	100	221	406	
Бразилия	611	24291	24752	577	24436	24962	532	24445	25077	521	24862	25394	872	25395	25267	
Бразилия	82	186	218	34	211	243	30	217	252	36	238	274	41	231	272	
Бразилия и Германия	155	168	261	158	115	272	109	137	308	72	223	366	100	221	406	
Бразилия	31	2	56	35	4	94	56	31	59	45	17	56	44	14	58	
Бразилия	362	2148	7808	966	229	1191	1029	231	1260	1089	237	1329	1127	261	1371	
Бразилия-Департамент	18	6	18	75	6	20	16	6	16	23	6	23	19	6	16	
Бразилия	36	17	53	31	17	48	13	16	26	17	25	43	17	25	43	
Бразилия	1949	9916	7876	2246	6484	8991	2241	3843	10361	2506	8524	11170	2708	9126	11024	
Бразилия	2919	2871	4796	2713	3410	6123	2941	3412	6353	3396	3261	6861	2903	2861	5086	

ЕМИСС и «Открытые данные»

Единая межведомственная информационно статистическая система (ЕМИСС) . В реестре - значительное количество показателей, так или иначе относящихся к миграции.

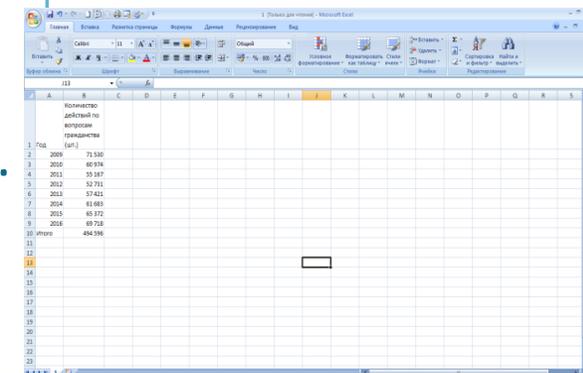
Перечень доступных в ЕМИСС показателей недостаточен для описания (и тем более анализа) миграционной ситуации, сами данные зачастую неактуальны. Много сведений не статистического характера.

Открытые данные.

Портал открытых данных создавался для «предоставления на едином информационном ресурсе *наиболее полных* перечней и сведений о наборах открытых данных, публикуемых органами государственной власти и органами местного самоуправления Российской Федерации, а также иных востребованных наборах открытых данных» (Пушкин В.М. 2016)

Однако, в отношении данных по миграции он явно не достиг своих целей.

(Подробнее см.: О.Чудиновских, Е. Донец. О новых технологиях и статистике миграции в России. Вопросы статистики, № 5 2018 г.)



The screenshot shows a Microsoft Excel spreadsheet with the following data:

Год	Количество действий по вопросам гражданства (шт.)
2009	71 588
2010	60 574
2011	35 987
2012	52 711
2013	57 421
2014	61 693
2015	69 372
2016	69 718
Итого	494 596

на скриншоте - открытые данные КД МИД о действиях по вопросам гражданства за 2009-2016 гг. <http://www.kdmid.ru/opendata/default1.aspx> доступ 6 июня 2018 г.

Новый феномен в информационном пространстве – big data.
Все чаще звучит идея, что они заменят традиционные данные по миграции и будут лишены их недостатков

Колоссальные объемы данных создаются в –результате звонков по мобильному телефону, отправки sms - сообщений, проведения платежей, покупок товаров и услуг, общения в социальных сетях, работы с поисковыми системами в интернете.....

Они могут отражать **текущее местоположение абонента или историю его перемещений**, т.к. накапливаются на серверах операторов, обслуживающих эти процессы.

С точки зрения изучения территориальных перемещений большие возможности открывает геолокация (определение реального географического **местоположения** электронного устройства, например радиопередатчика, сотового телефона или компьютера, подключённого к Интернету.)

Предполагается, что с помощью мобильных устройств можно изучать подвижность и поведение людей (M.Tiru 2014)

Агрегированные

данные сетей
мобильной
телефонной связи



данные
социальных сетей



Данные «умных»
(сканирующих и
передающих)
устройств



“Большие данные”–

1. большие по объему,
2. во многих случаях - неструктурированные наборы данных ,
3. часто доступные в режиме реального времени,
4. требующие применения новых методов и технологий для управления ими и их анализа....

Считается, что «официальной статистике крайне важно адаптироваться к меняющемуся информационному ландшафту» (Naylor, Swier, Williams. 2016)

Но некоторые исследователи миграции ожидают «революции в данных» , когда Большие данные станут использоваться для изучения миграции.....

Почему к БД особенный интерес с точки зрения измерения миграции?

Давняя история «претензий» к наличию и качеству традиционных данных по миграции:

Переписи - редки (и дороги)

Обследования - дорогие и не отражают всей популяции, трудно обеспечить охват мигрантов , т.к. миграция «редкое явление»

Регистры и разные административные источники - учитывают только декларируемые события

+ общий недостаток – статистика появляется с временным лагом , не оперативна

Большие данные – оперативны и объективны: отражают изменения плотности населения и маршруты движения , вне зависимости от желания самих людей и уведомлений

Могут компенсировать отсутствие в странах с низким уровнем жизни традиционных источников данных по миграции

Оперативно отражают переломные моменты в миграционных трендах

В теории - статистика на их основе должна обходиться дешевле (чем полученная с помощью традиционных источников)

Снижение бремени на население при сборе данных по сравнению с переписями, обследованиями и пр.

Ограничения больших данных при измерении пространственной мобильности населения

- ❖ Несколько мобильных устройств у одного человека
- ❖ Отсутствие переменных, требуемых в традиционной статистике миграции
- ❖ Доступ к данным (в большинстве случаев они корпоративные)
- ❖ Практически невозможно применить концепции, принятые в статистике миграции (в первую очередь – концепции места обычного / постоянного жительства)

Опыт применения больших данных для измерения мобильности людей

Исследования начались относительно недавно

Их не так много

Более понятные и удачные результаты на основе Больших данных операторов сотовой связи

Яндекс 2016: изучение внутригородской дислокации работы и дома и мобильности населения (но уже на основе другого источника – данных приложений «Яндекс-навигатор» и «Яндекс-Карты»), проект позволил выделить жилые, рабочие и смешанные районы города, а также оценить время и направления внутригородских поездок.

Правительство Москвы (проект осуществляется в настоящее время) - используются данные операторов сотовой связи для оценки размещения реального населения для более эффективного планирования градостроительной деятельности, оценки внутригородской маятниковой мобильности населения и пр.

Positium 2012 (Эстония) Оценка туристических потоков. Предложен алгоритм разработки информации операторов сотовой связи для изучения пространственной мобильности населения (точнее - абонентов) на примере трансграничных маятниковых поездок между Эстонией и Финляндией. Данные удалось структурировать по признаку времени, проведенного абонентами сотовой связи в «другой» стране, определить, откуда (конкретные населенные пункты) и куда перемещаются люди.

И пр.

О возможностях замеров миграции на основе больших данных

..... мы ищем здесь не точные цифры, а поворотные точки в миграционных потоках и масштабы одной волны миграции по сравнению с другой.

Цит по <https://www.forbes.com/sites/federicoguerrini/2014/08/08/using-big-data-to-understand-migrations/#2ac168803700>

Zagheni 2014.

IOM 2014 .Big Data and Migration

How data innovation can serve migration policymaking

От автоматизации к цифровизации (несколько риторических вопросов)



- ✓ Правильно ли все внимание уделять цифровизации, если не решены рутинные вопросы автоматизации важнейших процессов при производстве статистики населения и организации к ней доступа пользователей?
- ✓ Какова высшая цель цифровизации (и есть ли она)?
- ✓ Кто будет ее конечным бенефициаром и в какой форме будут существовать выгоды от цифровизации?



Заключительные замечания

1. Говоря о цифровизации в науке демографии в первую очередь нужно видеть потенциал новых технологий и новых видов данных для информационного обеспечения исследований и управленческих решений. Но мы исходим из предположения, что без решения проблем промежуточного этапа – автоматизации сбора и разработки статистики, обеспечения доступа широкого круга пользователей к агрегированным данным в электронном формате, нельзя говорить о реальной цифровизации. Равно как нельзя смещать весь фокус внимания только в сторону цифровизации.
2. В России основной комплекс проблем связан не только с технологиями сбора данных и производства статистики, но и с характером взаимодействия разных производителей данных между собой и всех производителей – с пользователями. Проблема старая и постоянно воспроизводится.
3. Потенциал новых видов данных - в первую очередь - больших данных как источника информации о населении и его пространственной мобильности велик и до конца не исследован. Им присущи объективные ограничения, которые надо принимать при разработке статистики. Они дополняют, но не замещают, традиционную статистику миграции, позволяя пролить свет на краткосрочные виды мобильности или такие ее формы, которые обычными методами измерить нельзя.
4. Какими бы мобильными ни стали люди будущего, вряд ли исчезнет феномен места постоянного жительства и потребность в измерении миграции на основе этой концепции. Сохранятся и традиционные источники данных, изменится только технология работы с информацией (накопления, хранения, разработки и пр.)
5. Следуя объективным трендам развития цифровых технологий нельзя забывать о человеческих и институциональных факторах. Их влияние на наличие и качество данных, а также доступ к статистике могут быть решающими. В России имеет место разрыв между декларируемыми целями и реальной ситуацией, прогнозировать развитие событий – трудно.

Цифровая реальность не отменяет и не решает всех проблем «традиционных» источников данных о населении и миграции, ими все равно придется заниматься.....

Спасибо за внимание