#### Институт демографии НИУ ВШЭ

Заседание Ученого Совета ИДЕМ 19 января 2012 г.

# База данных демографических показателей по регионам России и странам мира: концепция и принципы построения

Сороко Е.Л.

•Имеющиеся базы демографических данных: проблемы и недостатки

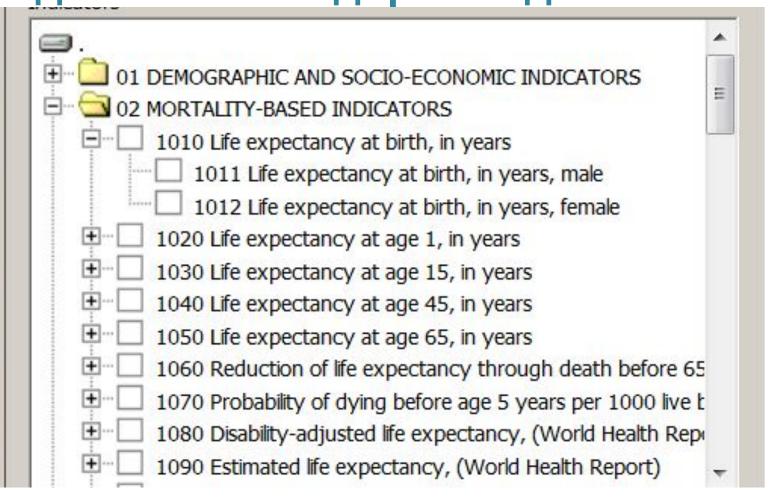
- •Номенклатура показателей
- •Описание алгоритмов и формул расчета
- •Описание исходных данных и их источников
- •Точность и формат
- •Интерфейс пользователя БД

# База данных демографических показателей ИДЕМ НИУ ВШЭ •Вопросы для обсуждения

- •Номенклатура показателей
- •Как организовать данные
- •Проблема множественности значений
- •Перечень стран мира
- •Каковы принципы метаописания массивов
- •Что такое жизненный цикл показателя
- •Что можно разместить на Демоскопе
- •Каково содержание Базы данных на сегодня

•Что такое демографический показатель?

•База данных ВОЗ Здоровье для всех



- •Что такое демографический показатель?
- •База данных Демографический взгляд

Индикаторы

<b>Население</b>	
Доля населения моложе 20 лет	
Доля населения старше 65 лет	
Доля населения старше 80 лет	

Смертность и продолжительность							
жизни							
Коэффициент младенческой смертности							
Средняя продолжительность предстоящей жизни при рождении для женщин							
Средняя продолжительность предстоящей жизни при рождении для мужчин							

### База данных демографических показателей ИДЕМ НИУ •Что такое демографический показатель?

- •Демографический показатель число?
- •Демографический показатель таблица?

Численность мужчин в возрасте 82 года Число прибывших из Таджикистана Число выехавших в Канаду

```
Показатель = {Собственно индикатор, Разрез (категории)}
```

#### •Что такое демографический показатель?

Резервы и эффект применения формулы
Показатель = {Собственно индикатор, Разрез (категории)}
Примеры:

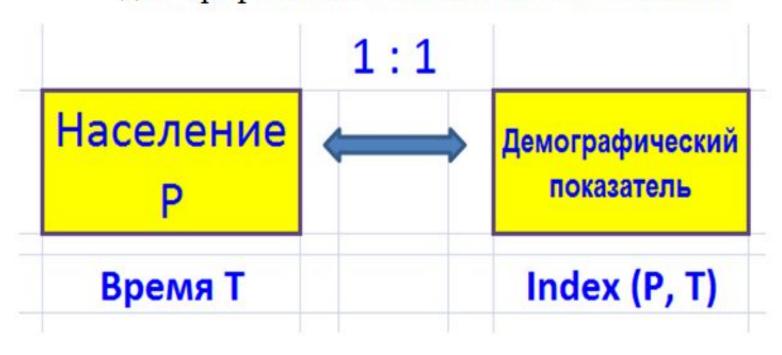
- •Показатели миграции в 3 раза
- •Таблица смертности показатель?

Направление миграции		Файл
Направление миграции	№ п	код
прибытие	1	in_m
иммигранты	1	in_m
отъезд	2	out_m
эмигранты	2	out_m
прибывшие	1	in_m
выбывшие	2	out_m
миграционный прирост	3	net_m
прибыло	1	in_m
выбыло	2	out_m
миграционный прирост	3	net_m
Число прибывших	1	in_m
Число выбывших	2	out_m
Миграционный прирост (убыль)	3	net_m

	Nº	код
mx	1	mx
qx	2	qx
ax	3	ax
lx	4	lx
dx	5	dx
Lx	6	Lx
Tx	7	Tx
ex	8	ex

# База данных демографических показателей ИДЕМ НИУ •Как можно решить проблему множественности значений показателя?

Рисунок 1. Традиционное соотношение демографического показателя и его значения



# База данных демографических показателей ИДЕМ НИУ •Как можно решить проблему множественности значений показателя?

Таблица 1. Значение показателя ожидаемой продолжительности жизни женщин в России по данным различных источников, лет\*

	F. M. (1) (2)			V. 6	
	Год	1997	1998	2005	2007
Источник	№ 11/11	5	6	7	8
BO3	1	72,96	73,27	72,40	-
ДЕ 2002	2	72,89	72,93		
ДЕ 2009	3	72,84	73,13	72,39	73,90
PRB	4			72,00	72,00

<sup>\* 1 —</sup> База данных «Здоровье для всех» Европейского Бюро ВОЗ [6], доступ 17.04.2011, 2, 3 — Демографический ежегодник России, 2002, 2009 [1], [2], 4 — Population Reference Bureau, World population Data Sheet, 2005, 2007 [9].

•Как можно решить проблему множественности значений показателя?

Примеры из «Населения России 2009»

#### Аборты:

- •Росстат
- •Минздравсоцразвития

#### Демографический прогноз

- •Росстат
- **•**00H
- •Бюро цензов

База данных демографических показателей ИДЕМ НИУ
•Как можно решить проблему
множественности значений показателя?

Вполне естественный вопрос: можно ли рассматривать значение показателя из одного из источников «правильным», «истинным», а другие, отличные от него – ошибочными?

# База данных демографических показателей ИДЕМ НИУ •Как можно решить проблему множественности значений показателя?

Дельта	CE2005-DS200	)2						
2 Hillian and He	1960	1970	1980	1990	1995	1999	2000	2001
Швеция	0	0	0	0	0	0	0	0
Австрия	0	0	0	- 51	- 99	- 100	- 100	- 93
Бельгия	35	18	12	0	0	0	0	1
Великобрит	<b>a</b> 5	31	- 15	- 313	- 583	- 825	- 866	- 895
Германия	- 11	458	- 17	69	0	0	0	5
Греция	6	0	1	- 4	180	345	358	368
Дания	2	- 5	0	1	0	0	0	0
Ирландия	- 3	7	12	8	0	- 1	4	11
Испания	- 226	- 1	53	- 1	121	216	241	405
Италия	0	0	0	0	0	0	0	- 976
Нидерландь	0	0	0	0	0	0	0	2
Норвегия	0	0	0	0	0	0	0	0
Португалия	0	0	0	2	3	- 2	- 5	- 6
Финляндия	0	0	0	0	0	0	0	0

### •Как можно регинть проблему

	множественности значений показателя?										
MH	ожесте	зенно	сти зн	ачени	и пока	азател	пя?				
Дельта DS2002	и Devision										
10 km 11 m 11 m 12 m 12 m 12 m 12 m 12 m 1	1960	1970	1980	1990	1995	1999	2000	2001			
Австрия	0	0	0	51	99	100	100	88			
Бельгия	- 35	- 18	- 12	0	0	0	0	- 1			
A 1,12,12,112,112,112,112,112,112,112,112											

Австрия	0	0	0	51	99	100	100	88
Бельгия	- 35	- 18	- 12	0	0	0	0	- 1
Германия	11	- 458	17	- 69	0	0	0	- 5
Греция	- 6	0	- 1	4	- 180	- 345	- 358	- 368

Германия	11	- 458	17	- 69	0	0	0	- 5
Греция	- 6	0	- 1	4	- 180	- 345	- 358	- 368
Дания	- 2	6	2	- 1	0	0	0	0
Ирпочния	2	7	10	0	0	1	4	11

Греция	- 6	0	- 1	4	- 180	- 345	- 358	- 368
Дания	- 2	6	2	- 1	0	0	0	0
Ирландия	3	- 7	- 12	- 8	0	1	- 4	- 11
Испания	226	1	- 53	1	- 121	- 216	- 291	- 455

Дания	- 2	6	2	- 1	0	0	0	0
Ирландия	3	- 7	- 12	- 8	0	1	- 4	- 11
Испания	226	1	- 53	1	- 121	- 216	- 291	- 455
Итапия	0	0	0	0	0	0	464	976

The state of the s								
Ирландия	3	- 7	- 12	- 8	0	1	- 4	- 11
Испания	226	1	- 53	1	- 121	- 216	- 291	- 455
Италия	0	0	0	0	0	0	464	976
Нидерланды	0	0	0	0	0	0	0	- 2
Норвегия	0	0	0	0	0	0	0	0

- 2

313

26

0

583

0

0

2

825

- 50

0

5

895

- 264

866

- 156

0

0

15

0

0

- 31

0

0

0

0

- 5

Португалия

Соединенное

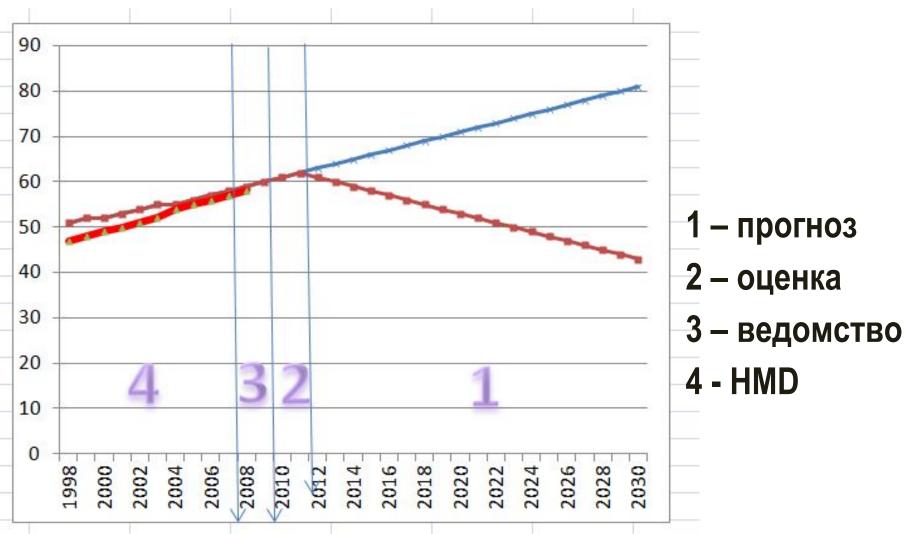
Финляндия

Швейцария

Франция

Швеция

# •Что такое жизненный цикл демографического показателя?

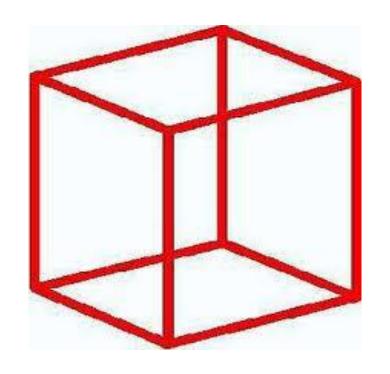


# База данных демографических показателей ИДЕМ НИУ ВШЭ Специфика демографической информации

обратная связь	•	дос	ка объ	явлен	ий	•	поис	CK	•	архив		n	ерево	ц trai	nslatio	n
										40 пр	омыш	ленно	разв	итых	стран	мира
		7. K	ффео	ициен	т сум	марно	жод й	сдаем	ости*,	1950-2	009					
															d	English
Страна								Го	ды							
Страна	1950	1960	1970	1980	1990	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Австралия	3,06	3,45	2,86	1,90	1,93	1,77	1,77	1,77	1,77	1,76	1,76	1,79	1,79	1,78	1,78	1,80
Австрия	2,09	2,70	2,29	1,65	1,46	1,34	1,36	1,33	1,39	1,38	1,42	1,41	1,41	1,38	1,41	1,39
Белоруссия	2,61	2,80	2,30	2,04	1,90	1,29	1,31	1,27	1,22	1,21	1,20	1,20	1,29	1,37	1,23	1,21
Бельгия	2,34	2,56	2,25	1,68	1,62	1,61	1,66	1,64	1,62	1,64	1,64	1,64	1,64	1,64	1,65	1,65
Болгария	2,94	2,32	2,17	2,05	1,82	1,23	1,30	1,24	1,21	1,23	1,29	1,31	1,38	1,42	1,48	1,57
Босния и Герцеговина	4,82	3,95	2,71	1,93	1,71	1,36	1,28	1,23	1,23	1,18	1,19	1,21	1,22	1,23	1,24	1,23
Великобритания	2,19	2,71	2,43	1,89	1,83	1,69	1,64	1,63	1,64	1,71	1,76	1,78	1,84	1,66	1,96	1,83
Венгрия	2,60	2,02	1,98	1,91	1,87	1,28	1,32	1,31	1,30	1,27	1,28	1,31	1,34	1,32	1,33	1,28
Германия	2,10	2,37	2,03	1,56	1,45	1,36	1,38	1,35	1,31	1,34	1,36	1,34	1,33	1,37	1,38	1,37
Греция	2,27	2,21	2,40	2,23	1,39	1,24	1,27	1,25	1,27	1,28	1,30	1,33	1,40	1,41	1,36	1,34
Дания	2,58	2,54	1,95	1,55	1,67	1,74	1,77	1,75	1,72	1,76	1,78	1,80	1,85	1,85	1,89	1,84
Ирландия	3,36	3,78	3,85	3,24	2,11	1,91	1,90	1,94	1,97	1,98	1,95	1,88	1,90	2,03	2,10	2,07
Испания	2,48	2,77	2,88	2,20	1,36	1,20	1,24	1,25	1,26	1,31	1,33	1,35	1,38	1,40	1,40	1,44
Италия	2,50	2,41	2,43	1,64	1,33	1,22	1,24	1,25	1,26	1,29	1,33	1,32	1,35	1,29	1,30	1,39
Канада	3,37	3,80	2,33	1,77	1,79	1,60	1,49	1,60	1,50	1,53	1,53	1,54	1,59	1,66	1,60	1,53
Латвия	2,00	1,94	2,01	1,90	2,00	1,18	1,24	1,21	1,23	1,29	1,24	1,31	1,35	1,41	1,45	1,32

http://demoscope.ru/weekly/app/app4007.php

### База данных демографических показателей ИДЕМ НИУ ВШЭ Специфика демографической информации



Data cube – информационный массив Может быть 1-, 2-, 3-, 4-, 5-... мерным Это – прямоугольник (2) или параллелепипед (3,4,...)

#### База данных демографических показателей ИДЕМ НИУ ВШЭ Как устроен информационный массив

- 1. Метаописание
- 2. Информационный массив

			Meta						
1	1	название показателя	Численность населения на начало года						
1	2	название информационного массива Численность населения на начало года по типу поселений РФ, 1							
1	3	№ показателя п/п	14						
1	4	код показателя	POP						
1	5	размерность информационного массива	2						
1	100	число переменных по вертикали	1						
1	111	название категории 1	Год						
1	112	№ категории 1 п/п	2						
1	113	код категории 1	YEAR						
1	114	Число строк категории 1	30						
1	200	число переменных по горизонтали	1						

- •Этапы подготовки информационного массива
- •Подготовка метаописания
- •Форматирование массива значений показателя
- •Проверка кодирования
- •Сохранение в виде текстового файла
- •Загрузка на сервер
- •Тестирование показателя

### База данных демографических показателей ИДЕМ НИУ ВШЭ База данных в интернете

http://db.demoscope.ru/bd\_devel.php

Виды страниц и разделы

- •Страницы для разработчика
- •Образцы информационных массивов
- •Документация
- •Нормативно-справочная информация
- •Список участников

#### База данных демографических показателей ИДЕМ НИУ ВШЭ Организация работ по разработке БД

- •Список показателей
- •Список источников данных
- •Список информационных массивов
- •Описание источников данных
- •Подготовка информационных массивов
- •Поддержание нормативно-справочной информации
- •Ведение сайта Базы данных
- •Отчетность по гранту Учитель-Ученики

- •Принципы и структура метаописания
  - •Максимальная детализация описания ИМ
  - •Гибкость метаописания
  - •Точность, гарантирование отсутствия неоднозначности

Как сделать правильно метаописание? <u>Инструкция по подготовке</u> <u>информационных массивов</u>

#### База данных демографических показателей ИДЕМ НИУ ВШЭ Зачем нужны коды?

- •Устранение синонимии
- •Выявление ошибок

,,,,,,,,,,,,,,,,,	VA.5	Страна											
№ n/n	код		1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
74	BEL	Беларусь									<mark>9751000</mark>	9714000	9690000
3	BEL	Белоруссия	10092986	10045237	10019480	9990435	9950941	9898590	9849062	9800073	9751000		
#Н/Д	#н/д	Республика Беларусь	10092986	10045237	10019480	9990435	9950941	9898590	9849062	9800073	9751000	9714000	9690000
#H/Д	#Н/Д	Беларуссия	10092986	10045237	10019480	9990435	9950941	9898590	9849062	9800073	9751000	9714000	9690000

#### База данных демографических показателей ИДЕМ НИУ ВШЭ Зачем нужны коды?

- •Устранение синонимии
- •Выявление ошибок

demo_frat	e-Fertility	y rates by	y age
Last update	11-02-2011		
Extracted on	02-03-2011	13:43:16	
Source of data	Eurostat		
GEO/TIME	2009	2008	2007
BE	1,83988	1,86323	1,81621
BG	1,56789	1,48053	1,41778
CZ	1,49239	1,49743	1,43662
DK	1,83868	1,88764	1,84230
DE	1,35691	1,37865	1,36875
DE_TOT	1,35691	1,37865	1,36875
EE	1,62218	1,65281	1,63133
IE	2,06776	2,09607	2,01031
GR	1,52200	1,50855	1,41278

# База данных демографических показателей ИДЕМ НИУ ВШЭ Участники разработки Базы данных <a href="http://db.demoscope.ru/bd\_parti.php">http://db.demoscope.ru/bd\_parti.php</a>

#### Список участников

ФИО	Раздел Проекта
Сороко Е.Л.	Разработка концепции и основных функций Базы данных, форматов информационных массивов, обучение персонала разработчиков
Абылкаликов С.И.	Анализ и подготовка информационных массивов демографических показателей по миграции населения Российской Федерации и регионов России
Аминов И.Г.	Анализ демографических показателей Центральной базы статистических данных Росстата, подготовка информационных массивов демографических показателей по России и ее регионам
Винник М.В.	Анализ и подготовка информационных массивов демографических показателей по Российской Федерации и регионам России из Демографических ежегодников Росстата
Данилова И.А.	Проверка форматирования подготовленных массивов демографических показателей, формирование и пополнение НСИ, проверка правильности кодирования значений признаков социально-демографических групп, проверка корректности мета-описания данных.
Двинянин Е.А.	Анализ демографических показателей базы данных по странам мира Национального института демографических исследований (INED), подготовка информационных массивов по странам мира

#### Еще вопросы для обсуждения и принятия

Какие источники являются обязательными, наиболее важными, авторитетными и имеющие максимальный приоритет?

- Росстат
  - •Центральная база статистических данных
  - •Сайт Росстата
  - •Демографический ежегодник
- •Евростат
- •Базы данных ООН, ВОЗ
- •Национальные статистические службы
- Human mortality database

#### База данных демографических показателей ИДЕМ НИУ Еще вопросы для ВВ уждения и принятия

Можно ли создать Базу данных безгантий источников демографических данных? Что должен включать в себя такой анализ, как описать его результаты? Какие есть результаты анализа источников данных? Можно ли эти результаты разместить на Демоскопе? <a href="http://db.demoscope.ru/bd\_sources\_di.php">http://db.demoscope.ru/bd\_sources\_di.php</a>

#### Источники демографической информции

Источник	Описание					
Демографический ежегодник России	Статистический сборник. Федеральная служба государственной статистики Российской Федерации					
Евростат	База данных Евростата					
ЦБСД Росстата	Центральная база статистических данных. Федеральная служба государственной статистики Российской Федерации					
Численность и миграция РФ	Статистический бюллетень "Численность и миграция населения Российской Федерации"					
Statistics Estonia	Сайт статистической службы Эстонии					
Statistics Sweden	Сайт статистической службы Швеции					
WHO Health for all database	База данных "Здоровье для всех" Всемирной организации здравоохранения					
World Fertility Patterns	Доклад ООН "Мировые тренды рождаемости" (World Fertility Patterns)					
World Urbanization Prospects	Доклад ООН "Перспективы мировой урбанизации" (World Urbanization Prospects)					

Еще вопросы для обсуждения и принятия решений

Должны ли в Базе данных содержаться материалы переписей населения, РиДМиЖ, микропереписи 1994 г, социальнодемографических обследований, результатов табуляции индивидуальных микроданных из других источников? Должна ли База данных содержать сведения, полученные не на основе официальной, регулярной, всеобщей статистики?

### База данных демографических показателей ИДЕМ НИУ ВШЭ Еще вопросы для обсуждения и принятия решений

Как организовать обновление информационного массива? Варианты

- •Полная замена старого массива новым
- •Только за последние 10 ? Лет
- •Только для тех категорий, где имеются отличия

#### База данных демографических показателей ИДЕМ НИУ ВШЭ Еще вопросы для обсуждения и принятия решений

Что делать с когортными показателями?

#### Варианты

- •Пока: разные названия для одного показателя для когорты и для периода. Пример: коэффициент суммарной рождаемости и итоговая рождаемость
- •Дополнительная категория: тип показателя (периодный / когортный)

#### Еще вопросы для обсуждения и принятия решений

- Какие данные НЕ должны содержаться в Базе данных?
- а) очень большие объемы массивов (HMD, HFD),
- б) информация, находящаяся только в одном источнике,
- в) смертность по причинам смерти (подробная, с ICD-кодами и указанием пересмотра)
- г) таблицы, например, переписей населения, содержащие несколько различных сложно скомпонованных показателей в одной таблице,
- д) «экзотические» показатели
- е) сомнительные с точки зрения авторских прав.

#### База данных демографических показателей ИДЕМ НИУ ВШЭ Еще вопросы для обсуждения и принятия

Где должна проходить граница между «чисто демографическими» или собственно демографическими показателями и другими, относящимися к социально-экономической сфере? Такой вопрос неизбежно возникает при рассмотрении таких источников, как например, OECD Family database (<a href="http://www.oecd.org/document/4/0.3746.en">http://www.oecd.org/document/4/0.3746.en</a> 2649 37419 37836996 1 1 1 37419. 00.html). Среди показателей этой базы данных

- •Время поездок на работу,
- •Обеспеченность детскими дошкольными учреждениями,
- •Неработающая и не учащаяся молодежь,
- •Уровень безработицы (пример из НР2009) и т.д.

### База данных демографических показателей ИДЕМ НИУ ВШЭ Еще вопросы для обсуждения и принятия решений

Нужен ли Базе данных интеллект? Зачем?

- •Автоматическая «синхронизация» разных единиц измерения показателей из разных источников в случае их несовпадения,
- •Учет «спроса» на различные показатели,
- •Учет наиболее востребованных разрезов и категорий,
- •«Память» предпочтений конкретного пользователя при последнем запросе к Базе данных
- •Обнаружение и предупреждение о возможных ошибках

#### База данных демографических показателей ИДЕМ НИУ ВШЭ Еще вопросы для обсуждения и принятия

Еще вопросы для обсуждения и принятия решений

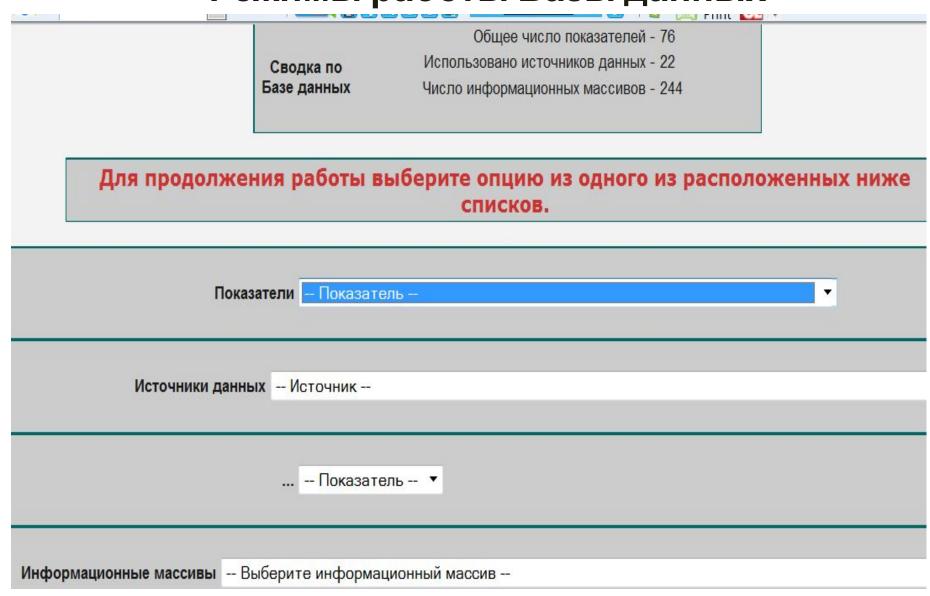
Какой перечень стран мира должен храниться в Базе данных?

На сегодня: все страны, которые содержатся в других базах данных (BO3, INED, OOH, Евростата), включая исторические (СССР, Югославия, Чехословакия, ...)

### База данных демографических показателей ИДЕМ НИУ ВШЭ Еще вопросы для обсуждения и принятия решений

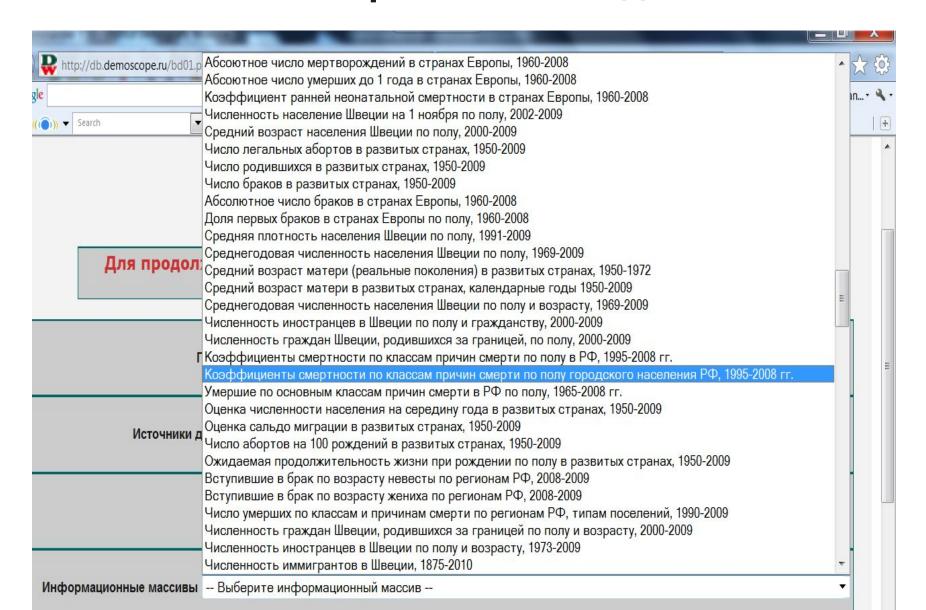
Что делать с ошибками?

Какие типы ошибок встречаются в исходных базах данных?



400	Среднегодовая численность населения				
	Младенческая смертность				
b.demoscope.ru/bd01.php	Число браков				
	Число разводов				
	Число абортов на 100 рождений				
- Go ▼ •	Коэффициент младенческой смертности по причинам смерти				
	Коэффициент мертворождаемости				
	Коэффициент неонатальной смертности				
Сво	Коэффициент перинатальной смертности				
	Внебрачная рождаемость				
	Средний возраст женщин при вступления в первый брак				
	Итоговая рождаемость женщин				
	Средний возраст матери при рождении ребенка				
Лля продолжения г	Демографическая нагрузка Общий коэффициент прироста населения				
для продолжения р	Общий коэффициент прироста населения				
	Национальный состав мигрантов				
	Гражданство международных мигрантов				
	Миграционный прирост населения				
	Коэффициент миграционного прироста				
Показатели	Число внебрачных рождений				
	Прогноз численности населения				
	Число мертворождений				
	Коэффициент ранней неонатальной смертности				
Источники данных И	Средний возраст населения				
-	Число легальных абортов				
	Доля первых браков				
	Плотность населения				
	Численность иностранцев				
•••	Численность граждан, родившихся за границей				
	Вступившие в брак по возрасту невесты				

ALC: ALC:	Источник
	База данных Евростата
noscope.ru/bd01.php	Демографический ежегодник России
	Демоскоп Weekly. Приложения.
	Росстат. Публикации. Демографический ежегодник России
▼ <b>③</b> Go	Центральная База Статистических Данных Росстата
	Statistics Bureau, Japan
	Statistics Estonia
	Population Reference Bureau. World Population Data Sheets
6	"Ежегодный демографический доклад ""Население России"" "
	INED database
	Росстат. Публикации. Демографический ежегодник России 2009
	Population Division, Departm. of Econ.and Social Affairs, UN Secretariat, World Population Prospects: 2008 Revision
77 PROFORWOUL	Statistics Sweden
ля продолжени	Численность и миграция населения Российской Федерации
	World Urbanization Prospects
	Центральная статистическая база данных
	Демографический ежегодник
	Численность и миграция населения Российской Федерации в 2008 году
Показат	Численность и миграция населения Российской Федерации в 2009 году
	Численность и миграция населения Российской Федерации в 2010 году
	World Fertility Patterns 2009
	Европейская база данных Здоровье для всех (HFA-DB)
Источники данных	
источники данных	ИСТОЧНИК



```
ami_004.xls · Среднегодовая численность населения по типу поселения и регионам России, 1990-2008
```

dc\_014.xls • Среднегодовая численность населения по промышленно развитым странам, 1950-2009

dc\_027.xls • Среднегодовая численность населения по всем странам мира, 2001-2010

ust\_006.xls • Среднегодовая численность населения в странах Европы, 1960-2009

ust\_007.xls • Среднеговая численность населения в странах Европы по полу, 1960-2009

уum\_025.xls • Среднегодовая численность населения Швеции по полу, 1969-2009

уum\_026.xls • Среднегодовая численность населения Швеции по полу и возрасту, 1969-2009

dvi\_022.xls • Оценка численности населения на середину года в развитых странах, 1950-2009

ami\_033.xls • Численность городского и сельского населения по странам мира, 1950-2050

Дата последней модификации информационных массивов: 13.10.2011.

База данных содержит сведения по 263 странам мира.

Данные массивы получены из следующих источников:

#### Источники

- 1 База данных Евростата
- 2 · Демоскоп Weekly. Приложения.
- 3 Центральная База Статистических Данных Росстата
- 4 Population Reference Bureau. World Population Data Sheets
- 5 · INED database

#### База данных демографических показателей ИДЕМ НИУ ВШЭ Состояние Базы данных

Где найти более подробную информацию о Базе данных?

Все материалы размещены на сайте Базы данных.

В данный момент это страница разработчика

http://db.demoscope.ru/bd\_devel.php,

см. также

http://db.demoscope.ru/bd\_parti\_lab.php,

http://db.demoscope.ru/doc/soroko\_lomonosov\_2011\_.doc,

http://db.demoscope.ru/doc/soroko\_hsehse\_2012.hse\_2012.pdf.