

ТЕХНОЛОГИИ BIG DATA: ПРОБЛЕМЫ КАЧЕСТВА ДАННЫХ МЕДИЦИНСКОЙ СТАТИСТИКИ



Обработка, анализ, моделирование

ПРОБЛЕМА

Настоящее

1. Множество медицинских, экологических, социально-экономических данных
2. Каждый день новые миллионы записей историй болезни, результаты клинических испытаний, показания экологического мониторинга
3. Каждый день новые связи между данными
4. Мы слишком мало знаем о заболеваниях, причинах, факторах влияния и лечении. Данные не стали знанием

Будущее

Мы можем узнать намного больше, если будем анализировать эти данные, не отдельно, а в совокупности и во взаимосвязях

BIG DATA

Зачем?

1. Повышение эффективности управления здравоохранением
2. Диагностирование болезней на ранних стадиях
3. Прогнозирование течения болезни
4. Персонализация лечения
5. Выявление ошибок лечения
6. Повышение качества и эффективности лечения
7. Профилактика заболеваний
8. Сокращение расходов на лечение и клинические испытания лекарственных препаратов

ЧТО ТАКОЕ BIG DATA ?

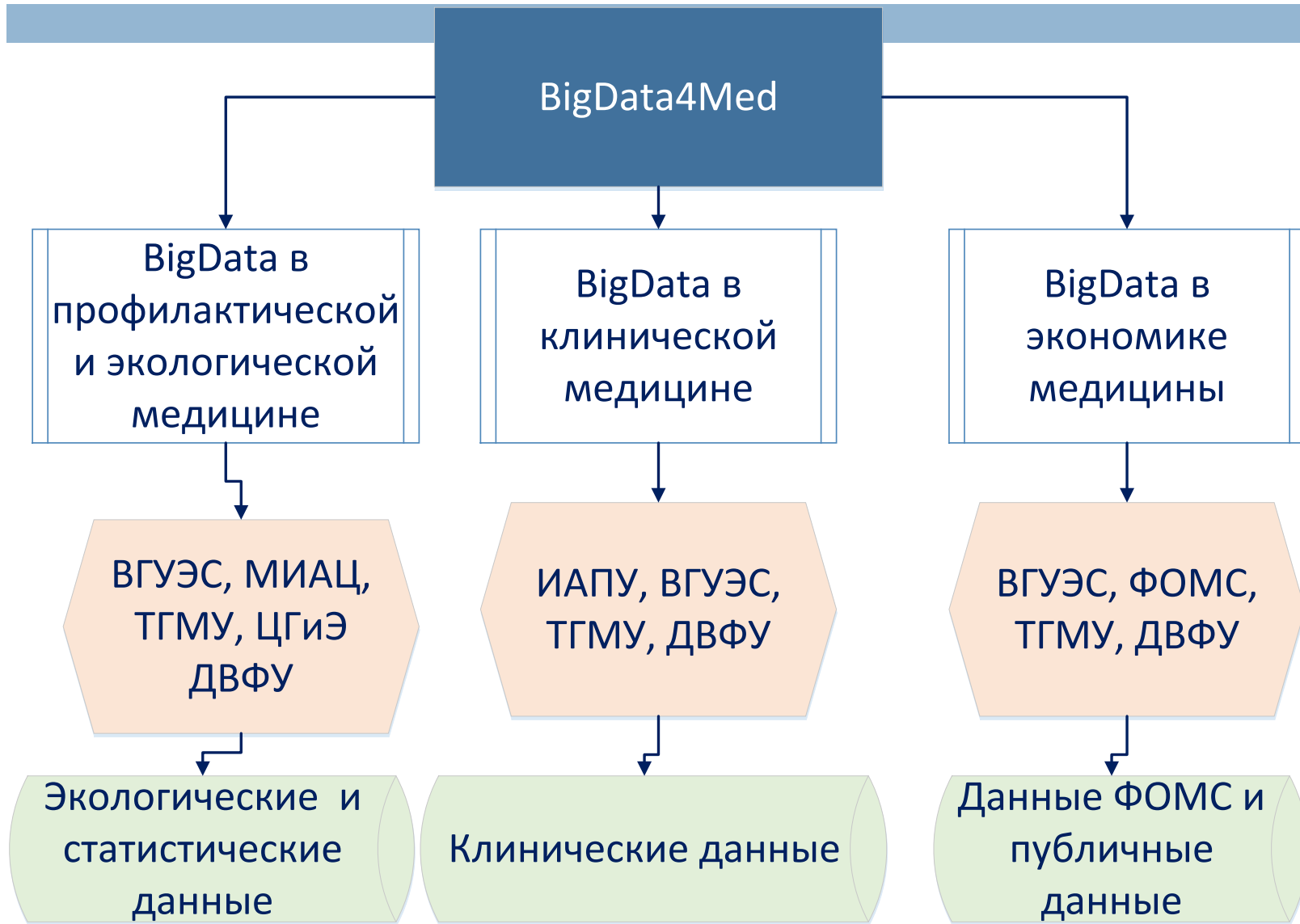
6 V

Big Data – это большие объемы высокоскоростных, сложных и переменных данных, которые требуют расширенных методов и технологий сбора, хранения, распространения, управления и анализа

value, volume, velocity, variety, **veracity**, and Variability

Veracity – качество больших данных сомнительно

ПРОЕКТ BigData4Med



ПРОЕКТ BigDataMed

Участники проекта

Руководитель проекта: член–корр. РАН Б.И. Гельцер

1. МИАЦ ПК (владелец данных и анализ)
2. ЦГиЭ (владелец данных и анализ)
3. ВГУЭС (ИТ–направление, обработка и анализ)
4. ТГМУ (медицина)
5. ИАПУ (Интеллектуальные медицинские системы)
6. Школа биомедицины ДВФУ (медицина и экология)

ДААННЫЕ МЕДИЦИНСКОЙ СТАТИСТИКИ

Проблемы данных

Данные Medstat

Отчет за 1 год \Rightarrow Невозможен анализ динамики

Отсутствует
полноценная
формализация
(нет справочников)

Нет однозначной
идентификации
объекта в рамках
периода

Постоянная смена
отчетных форм и
атрибутов внутри
формы

ДААННЫЕ МЕДИЦИНСКОЙ СТАТИСТИКИ

Синтаксические проблемы данных

1. Частичный ввод кириллицы латинским алфавитом
2. Орфографические ошибки ручного ввода данных

ДААННЫЕ МЕДИЦИНСКОЙ СТАТИСТИКИ

Семантические проблемы данных

1. Отсутствие формализации несмотря на реляционную модель

- Данные хранятся как отчетная форма, а не как формализованный объект:

Вместо «участковый педиатр» – «из них:
педиатр участковый»

ДААННЫЕ МЕДИЦИНСКОЙ СТАТИСТИКИ

Семантические проблемы данных

1. Отсутствие формализации несмотря на реляционную модель

- Нет справочников, даже МКБ-10, должностей

2004–2007 – Стенокардия I20

2008 – Ишемическая болезнь сердца I20–I25, в том числе
стенокардия I20

2009 – 2014 – Ишемическая болезнь сердца I20–I25, из них
стенокардия I20, из нее нестабильная стенокардия I20.0

ДААННЫЕ МЕДИЦИНСКОЙ СТАТИСТИКИ

Семантические проблемы данных

2. Ежегодные изменения идентификационного кода и названия объекта

- До 2011 «отоларингологи», с 2012– «оториноларингологи»
- До 2007 «наркологи», с 2008– «психиатры–наркологи»
- До 2005 «участковые терапевты цеховых участков», 2006–2007 «участ.терапевты цехов.участков», 2008–2011 «терапевты цехов.участков», 2012– н.вр. «терапевты цеховых участков»

ДАННЫЕ МЕДИЦИНСКОЙ СТАТИСТИКИ

Семантические проблемы данных

2. Ежегодные изменения идентификационного кода и названия объекта – слияние объектов

Год	Код	Название	Ставки	Физ. лица
2005	030100805	Детская поликлиника N 1	130,5	66
2006	–	–		
2012	030160806	Владивостокская поликлиника дет. №1	168,5	98
2005	030200805	Детская поликлиника N 15	115,25	59
2006	030200806	Поликлиника дет. №15 г.Вл-к	180,5	91
2012	030200806	Владивостокская поликлиника дет. №5	172,5	84

ДААННЫЕ МЕДИЦИНСКОЙ СТАТИСТИКИ

Семантические проблемы данных

3. Деление объектов и атрибутов объектов

- До 2014 «прочие врачи», с 2014 «врачи по медицинской профилактике», «врачи по медицинской реабилитации» и т.д.
- До 2008 физ. лица мед работников всего, с 2009 – физ. лица в амбулаториях и стационарах

ДААННЫЕ МЕДИЦИНСКОЙ СТАТИСТИКИ

Семантические проблемы данных

3. Включение объектов

- Объект «психиатры» включает объект «из них участковые»?
- Объект «психиатры–наркологи» включает объект «из них психиатры–наркологи участковые»
- «Педиатры всего» включает объекты «из них участковые педиатры» и «педиатры городские»?

ДААННЫЕ МЕДИЦИНСКОЙ СТАТИСТИКИ

Решение проблем

1. Замена латиницы на кириллицу
2. Выбор эталона и условного эталона
 - МКБ-10 – эталон
 - Отчетные формы из «Консультант» – условный эталон
 - Использование технологий MS Data Quality Server и Elastic Engine (соответствие эталону)
3. Поэтапная замена данных на эталон

ДААННЫЕ МЕДИЦИНСКОЙ СТАТИСТИКИ

Решение проблем

4. Определение отношений вложенности объектов на основе слов «в том числе», «из них», «включая»
5. Определение вложенности на основании других факторов

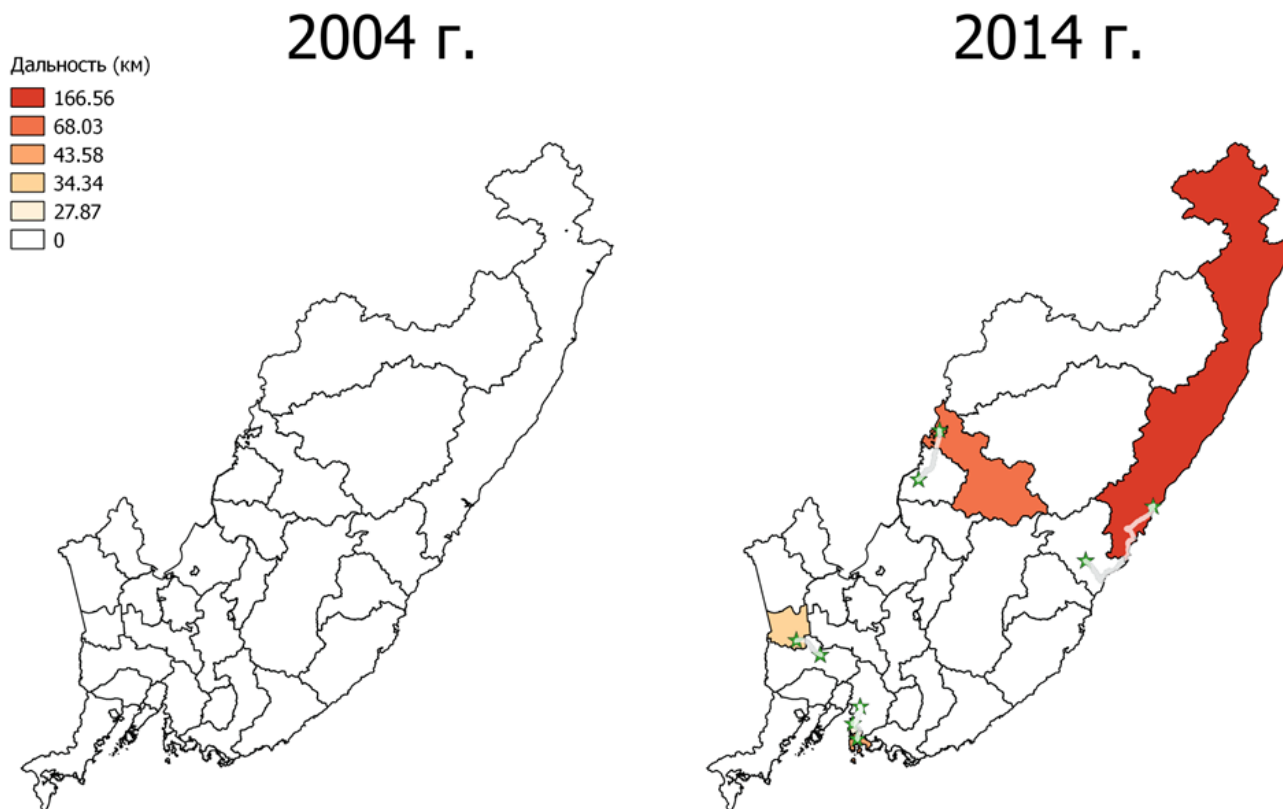
ДААННЫЕ МЕДИЦИНСКОЙ СТАТИСТИКИ

Решение проблем

6. Использование стоп-листа для корректировки названия объектов
7. Слежение за объектом: на основе атрибутов объектов автоматическое определение наиболее вероятных вариантов объединения, переименования и изменения идентификационного кода
8. Использование линейной регрессии для моделирования «назад»

Некоторые результаты анализа (R-studio и Qgis)

Показатели доступности фтизиатрической помощи
Территориальная доступность



Некоторые результаты анализа

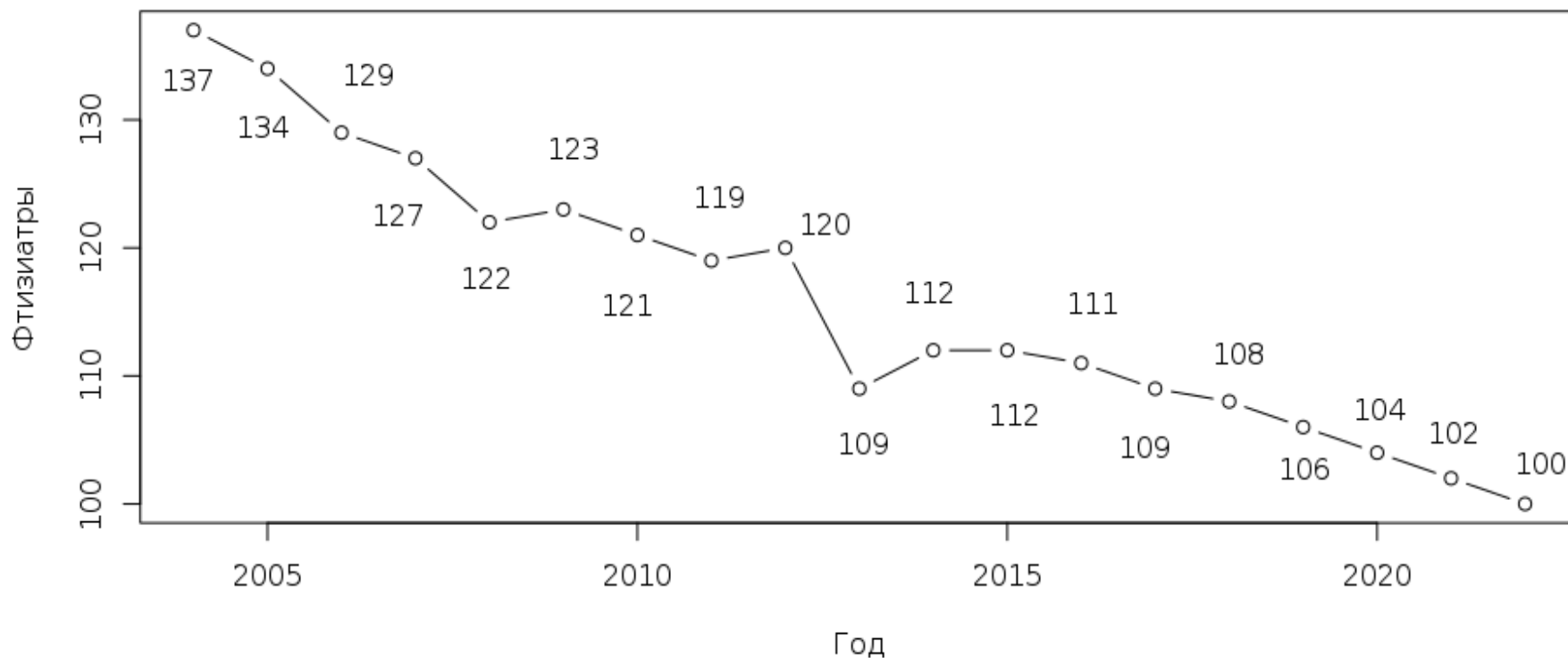
Показатели доступности фтизиатрической помощи

Коэффициент потребности Приморского края во фтизиатрической службе

Территория	Территориальная удаленность, км	Численность населения	Коэффициент потребности, усл. ед.
г. Большой камень	27,88	38637	0,14
г. Дельнереченск и Дальнереченский р-н	68,04	36641	0,33
Октябрьский р-н	34,34	28238	0,13
Тернейский р-н	166,57	11633	0,26
г. Фокино	43,58	23167	0,13

Некоторые результаты анализа

Показатели доступности фтизиатрической помощи
Кадровая доступность



Некоторые результаты анализа

Показатели доступности фтизиатрической помощи

Инфраструктурная доступность

Показатели доступности	Заболеваемость	Распространенность	Клиническое излечение	Смертность
Средняя территориальная доступность	-0,54	-0,2	-0,01	-0,83
Число коек на 100 тыс. населения	-0,11	0,04	0,12	0,9
Количество пациентов на 1 врача в стационаре	0,62	0,23	-0,18	0,75
Количество посещений на 1 врача в амбулаториях	-0,59	0,22	0,18	-0,42

Некоторые результаты анализа

Оценка кадрового ресурса ФП

Кадровый ресурс ФП	Население
Штатные должности	-0,45
Физ. лица фтизиатров	0,93
Амбулаторные штатные должности	0,68
Амбулаторные фтизиатры	0,39
Стационарные должности	-0,61
Стационарные фтизиатры	0,59

Некоторые результаты анализа

Оценка кадрового ресурса Врачей в целом

Характеристика	Средний ежегодный прирост/Спад
Выделенные ставки врачей	34,25
Занятые ставки врачей	-39
Физ. лица врачей	-103,9
Население	-11282

КОНТАКТЫ



ВГУЭС

Директор института ИТ ВГУЭС, д.т.н.,

Шахгельдян Карина Иосифовна,

+7-924-231-44-91

8-423-240-42-26

carinash@vvsu.ru

Skype: carinaish