



IV СЪЕЗД МЕДИЦИНСКИХ
СТАТИСТИКОВ ГОРОДА МОСКВЫ

**ЦИФРОВАЯ
СТАТИСТИКА**

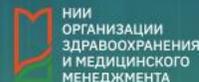
НОВЫЕ ЗАДАЧИ И ТРАЕКТОРИЯ
ДВИЖЕНИЯ

"Информационные источники для медицинской статистики, типы и особенности. Возможность использования бигдаты"

секция «От сбора к анализу: многообразные агрегаторы медицинской статистики»



Департамент
здравоохранения
города Москвы



НИИ
ОРГАНИЗАЦИИ
ЗДРАВООХРАНЕНИЯ
И МЕДИЦИНСКОГО
МЕНЕДЖМЕНТА

Базовые тезисы

- **Источники первичных данных и их обработка**
- **Типы и форматы данных, их свойства и потенциал для статистики**
- **Новые источники данных (медгаджеты, данные соц.сетей и др.)**
- **2 парадигмы обработки и управления агрегацией для аналитики**
- **Онтология данных в цифровом мире и «трансоотраслевая мультимодальность»)**
- **Социально-политический аспект кросс-аналитики по реальным первичным данным**

Первичная информация/данные в здравоохранении

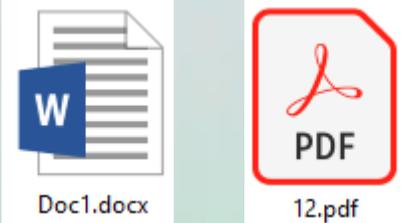
Возникновение данных

- Врачебные записи электронной медицинской карты
- Данные объективной диагностики (лабораторной и инструментальной)
- Данные по посещаемости (обращениям)
- Отчетные формы и реестры (вторичные данные)
- Демография (ЗАГС и др.)

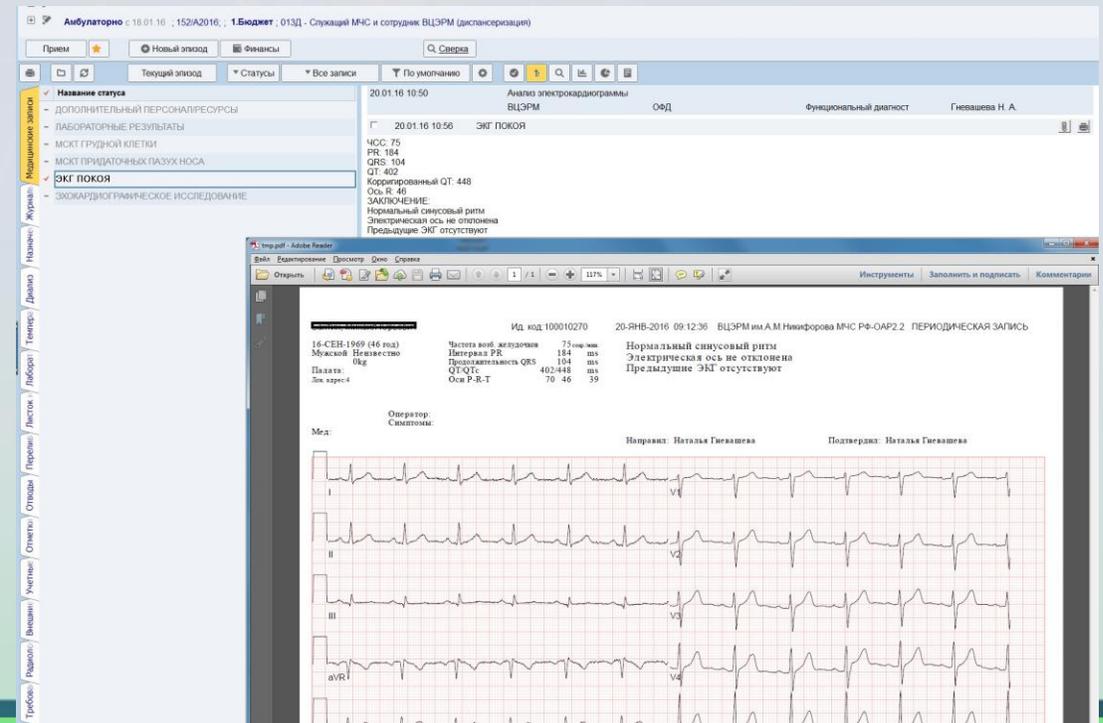
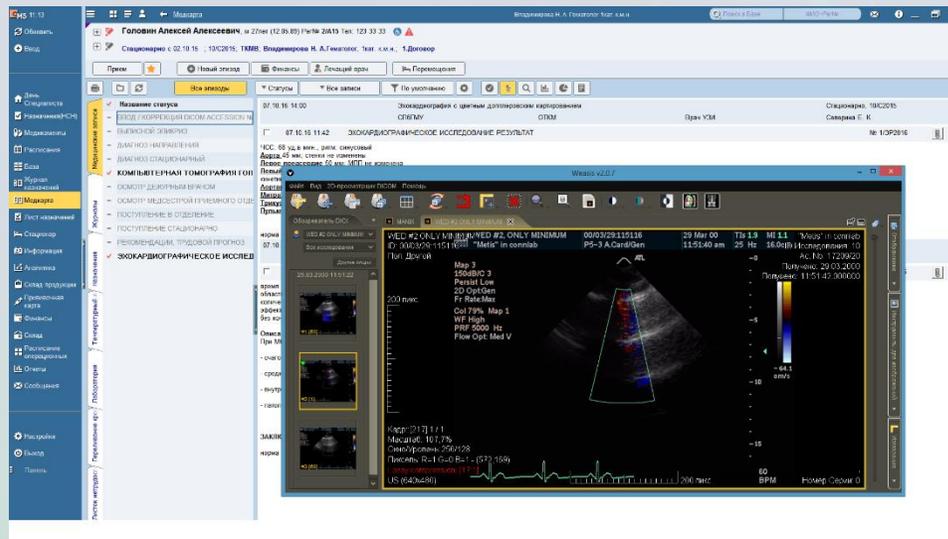
Обработка данных

- Сводные «таблицы» первичных регистрируемых данных
- Структурированные данные из ЭМК
- Автоматический анализ электронной медицинской карты
- Массив обработанных статистических форм и реестров

Типы структурирования данных

	Не структурированная	Структурированная
Бумажная		
Электронная		<pre data-bbox="1791 786 2244 1140"><?xml version="1.0" encoding="UTF-8"?> <xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"> <xs:element name="package" type="typeA"/> <!-- тип typeA --> <!-- тип typeA --> <xs:complexType name="typeA"> <xs:sequence> <xs:element name="tag"> <xs:complexType> <xs:complexContent> <xs:restriction base="typeB"> <xs:choice> <xs:element name="tag" type="xs:string"/> </xs:choice> </xs:restriction> </xs:complexContent> </xs:complexType> </xs:element> </xs:sequence> </xs:complexType> <!-- тип typeB --> <xs:complexType name="typeB"> <xs:choice> <xs:element name="tag" type="xs:string"/> <xs:element name="tag" type="xs:time"/> </xs:choice> </xs:complexType> </xs:schema></pre>

Результаты объективных исследований и их отражение в электронной медицинской карте (ЭМК)



Два шага до мечты... (шаг 1)

- Электронный рецепт (ЭР) – цифровой аналог обычного бумажного рецепта, который имеет равную с ним юридическую силу. В начале 2018-го года вступили в силу поправки в федеральный закон [№61-ФЗ](#) (п. 53 ст. 4). С этого момента, с согласия пациента, рецепт на лекарственный препарат может существовать в форме электронного документа.
- Согласно [паспорту](#) национального проекта «Здравоохранение» (2018 г.), уже к концу 2023-го года электронные рецепты должны будут применяться во всех субъектах РФ.

Два шага до мечты... (шаг 2)

**Приказ МЗ РФ от 07.09.2020 г. № 947н
«Об утверждении Порядка
организации системы
документооборота в сфере охраны
здоровья в части ведения
медицинской документации в
форме электронных документов»**

Вступил в силу с 1 февраля 2021 г.

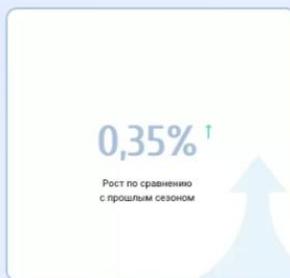
**Методические рекомендации «По
переходу на ведение медицинской
документации в форме электронных
документов»**

Утверждено 5 августа 2021 г.

ЕГИСЗ – Подсистема сбора информации о показателях системы здравоохранения и представления отчетности (что работает?)

- Автоматический сбор данных из подсистем Единой системы и иных государственных информационных систем о лицах, которым оказывается медицинская помощь, а также о лицах, в отношении которых проводятся медицинские экспертизы, медицинские осмотры и медицинские освидетельствования – **ЕГИСЗ, ВИМИС, РМИС (в Москве – ЕМИАС) ?**
- Ручной ввод данных в отчетные формы ?
- Автоматизированное формирование и предоставление сводной статистической и аналитической отчетности и ведение статистического наблюдения по отдельным формам статистического учета и отчетности?

2 встречных подхода (структурированные данные и/или извлечение слабо структурированных данных из ЭМК)



Год	2020	2021	2022	2023	Итого
Банки	123 ↑ Приход				
Банк 1	100 000	110 000	120 000	130 000	460 000
Банк 2	210 000	220 000	230 000	240 000	900 000
Банк 3	320 000	330 000	340 000	350 000	1 340 000
Банк 4	430 000	440 000	450 000	460 000	1 780 000
Итого	1 060 000	1 100 000	1 140 000	1 180 000	4 380 000

Сервисы сбора и извлечения данных из ЭМК

Сервис Webiomed.NLP

Отвечает за извлечение из неструктурированных ЭМК признаков в машинно-читаемом формате

2742
Поддерживаемых признаков

- ✓ До 80% клинически-значимой информации хранится в ЭМК в неструктурированных текстовых записях
- ✓ Webiomed автоматически извлекает из ЭМК нужные данные (признаки) с помощью NLP-сервиса
- ✓ Врачам не нужно заполнять специальные экранные формы с множеством полей или вести отдельные регистры и мониторинги.
- ✓ Вся нагрузка по извлечению информации из ЭМК Webiomed берет на себя.
- ✓ Этим мы экономим время врача на приеме, давая возможности больше заниматься пациентом

Сервис Webiomed.DataSet

Отвечает за централизованное хранение извлеченных и очищенных данных, пригодных для формирования наборов данных (data set) и работы моделей машинного обучения

4,6 млн.
Пациентов

36 млн.
Случаев лечения

146 млн.
Медицинских документов

462 млн.
Извлеченных признаков

✓ Социальные данные и анамнез

- ✓ Дата рождения, пол, регион проживания, социальная категория и т.д.
- ✓ Зарегистрированные диагнозы, рост, вес, окружность талии, курение
- ✓ Наследственность, история обращений

✓ Клинико-морфологические данные

- ✓ Данные инструментального и лабораторного обследования и т.д.
- ✓ Данные врачебных осмотров, протоколов хирургического лечения и т.д.
- ✓ Данные лекарственного лечения
- ✓ Данные скрининга, диспансеризации, анкетирования пациентов и т.д.

10

Ограничения и возможности

Помимо плохого структурирования и недостаточности данных существует проблема адекватности отражения в ЭМК фактической информации.

Сопоставление данных из статотчетов и ЭМК выявит много различий.

Главные проблемы с RWD на основе ЭМК



ИСТОЧНИКИ МЕДИЦИНСКОЙ ИНФОРМАЦИИ



Собранные со слов данные



Объективные данные



Данные с диагностического оборудования



ЭЛЕКТРОННАЯ МЕДИЦИНСКАЯ КАРТА



Ручное внесение данных в интерфейс ЭМК

Автоматическая передача данных от оборудования в МИС



ОСОБЕННОСТИ ВЕДЕНИЯ ЭМК

- **До 80% записей** – текстовые документы, неструктурированные данные
- **Чем более формализованная** экранная форма для внесения информации – **тем выше качество** и ниже скорость заполнения



Формализованный первичный осмотр

Порядка 500-900 полей, справочники значений, ср. время заполнения – 30-40 мин.



Первичный осмотр на основе шаблона

Порядка 12 полей, шаблоны готовых текстов, ср. время заполнения 3-7 мин.

- **Повторное использование** (копирование) данных из предыдущих документов
- **Отсутствие форматно-логического контроля** полноты и достоверности данных

Главный критерий ЭМК – удобство и скорость внесения информации

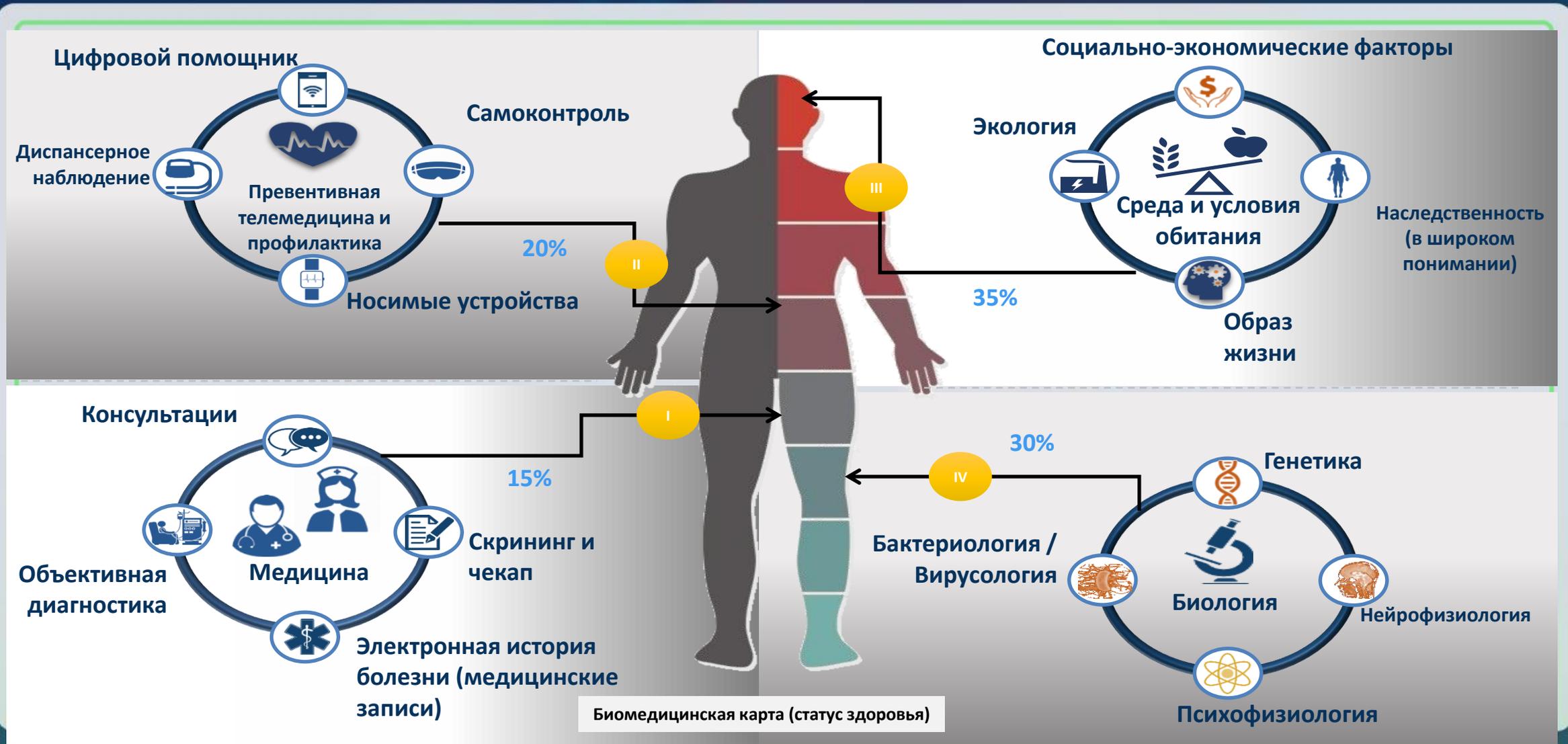
В целях статистики обычно не требуется 100% точность

Статистика должна браться напрямую из первичной медицинской документации (а точнее из событий и документов), а не из специально формируемых (часто вручную) статистических отчетов.

Это возможно и в ряде случаев происходит уже сейчас.

Не вся ЭМК – это «плохие» данные

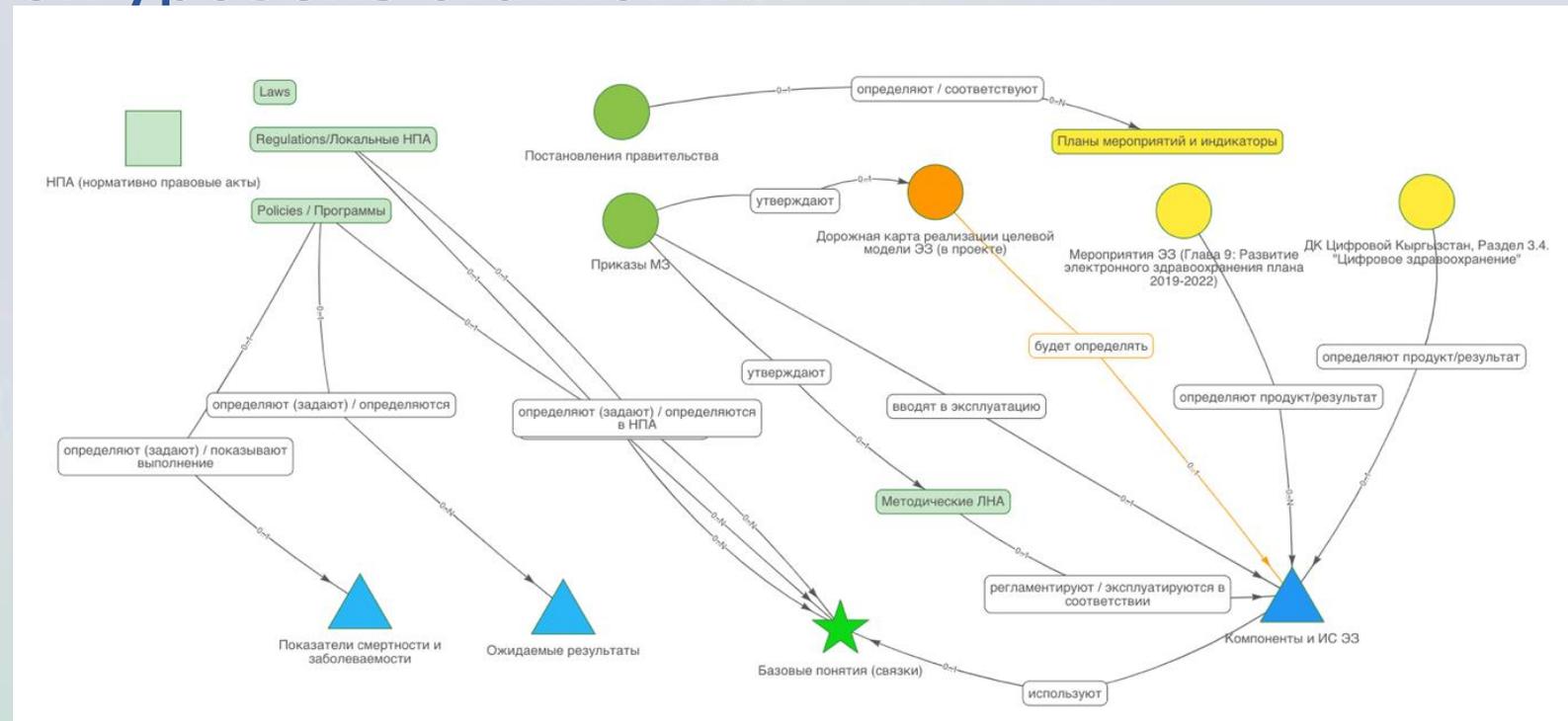




Онтологическая связанность данных, основа «управления знаниями» = другой уровень статистики

Цифровая трансформация управления требует нового качества работы с данными!

Данные должны детерминированы и описаны в онтологиях, тогда можно получить возможность их динамического структурирования и увидеть взаимосвязи.



пример взаимосвязки показателей национальных программ, планов мероприятий и мониторинга результатов на основе первичной информации.

Мультимодальность (определение)

Феномен, в широком смысле описывающий соединение нескольких модусов восприятия информации в процессе коммуникации.

Исследования мультимодальности ведутся параллельно в разных дисциплинарных полях и соответственно акцентируют внимание на разных аспектах этого феномена.

источник (статья): <https://www.hse.ru/ma/visual/articles/multimodality>

Трансдисциплинарность (определение)

- принцип рассмотрения любого предмета исследования вне рамок научных дисциплин как продуктов специализации
- взаимообмен данными и онтологическая связанность сущностей в различных отраслях и предметных областях

«Мультимодальная» и «Трансотраслевая» кросс-проверка (валидация)

- При наличии нескольких источников и многообразии типов данных в статистику становится возможной выявление новых зависимостей и трендов, на основе которых далее можно провести проверку одного источника в сравнении с другим и выявить искажения.
- Анализ больших данных (bigdata, озера данных и др.) средствами ИИ также может показать с достоверной точностью реальные цифры вне зависимости искажений в отчетности.

Социально-политический аспект...

Хороша ль, плоха ли весть, —
Докладай мне все как есть!
Лучше горькая, но правда,
Чем приятная, но лесть.

Леонид Филатов, "Про Федота-стрельца"

продолжение знаете...

Резюмируем:

1. Данные для статистики должны (и уже вполне могут) браться автоматически из первичной информации, а не вносятся руками в агрегированном виде. Вся агрегация только по формулам автоматически!!!
2. Данные могут «обогащаться» за счет «мультимодальных источников» и «трансдисциплинарных» подходов.
3. Цифровая трансформация в статистике может получить новый импульс с широким применением нейросетей (т.н. искусственного интеллекта) для обработки не структурированных «озер данных», но и онтологии не устарели.