

# Медицинские информационные системы

Особенности медицинских информационных систем. Особенности сбора и структурирования биомедицинских данных

Ланцберг Анна Вильямовна  
К.т.н., доцент кафедры ИУБ (Компьютерные системы и сети)  
[lantsberg\\_av@bmstu.ru](mailto:lantsberg_av@bmstu.ru)  
Каб. 801 ГК

# Важная особенность МИС:

Сбор, хранение и обработка разнородных, разноформатных, разнотипных **слабоструктурированных (полуструктурированных или плохоструктурированных) данных:**

- МИС могут включать большое количество различных показателей, классификация которых не разработана.
- Медицинские данные зачастую имеют неизвестную структуру, например, мультимедийные данные или документы со структурированным текстом
- Медицинские данные подвергаются частому изменению

## Виды данных в МИС:

- I Количественные показатели состояния организма (непрерывные и дискретные величины)
- II Вербальные описания (так называемые, качественные показатели)
- III Данные о биосигналах (например, электрокардиограмма, электроэнцефалограмма)
- IV Фото, видео, аудио, графическая информация (например, статические (рентгенограмма конечностей, органов дыхания) и динамические (рентгенограмма ЖКТ с использованием контрастной жидкости) картины органов человека или всего его тела)

## Процедуры сбора, хранения и обработки информации должны учитывать:

- Совокупность методов, используемых для кодирования информации
- Совокупность методов, используемых для предварительной обработки данных и их последующего анализа
- Совокупность стандартов, используемых для хранения и передачи данных в выбранной предметной области
- Модели данных, используемые для хранения информации

# Понятия, используемые при оценке медико-биологических данных

**Признак** – характеристика пациента (или явления), которая может иметь только два значения: наличие или его отсутствие (например, наличие болей, лихорадка, покраснение кожных покровов)

**Параметр** – это величина, характеризующая свойство процесса, явления или системы в абсолютных или относительных величинах (например, показатели температуры, артериального давления, изменение интервалов между зубцами на ЭКГ)

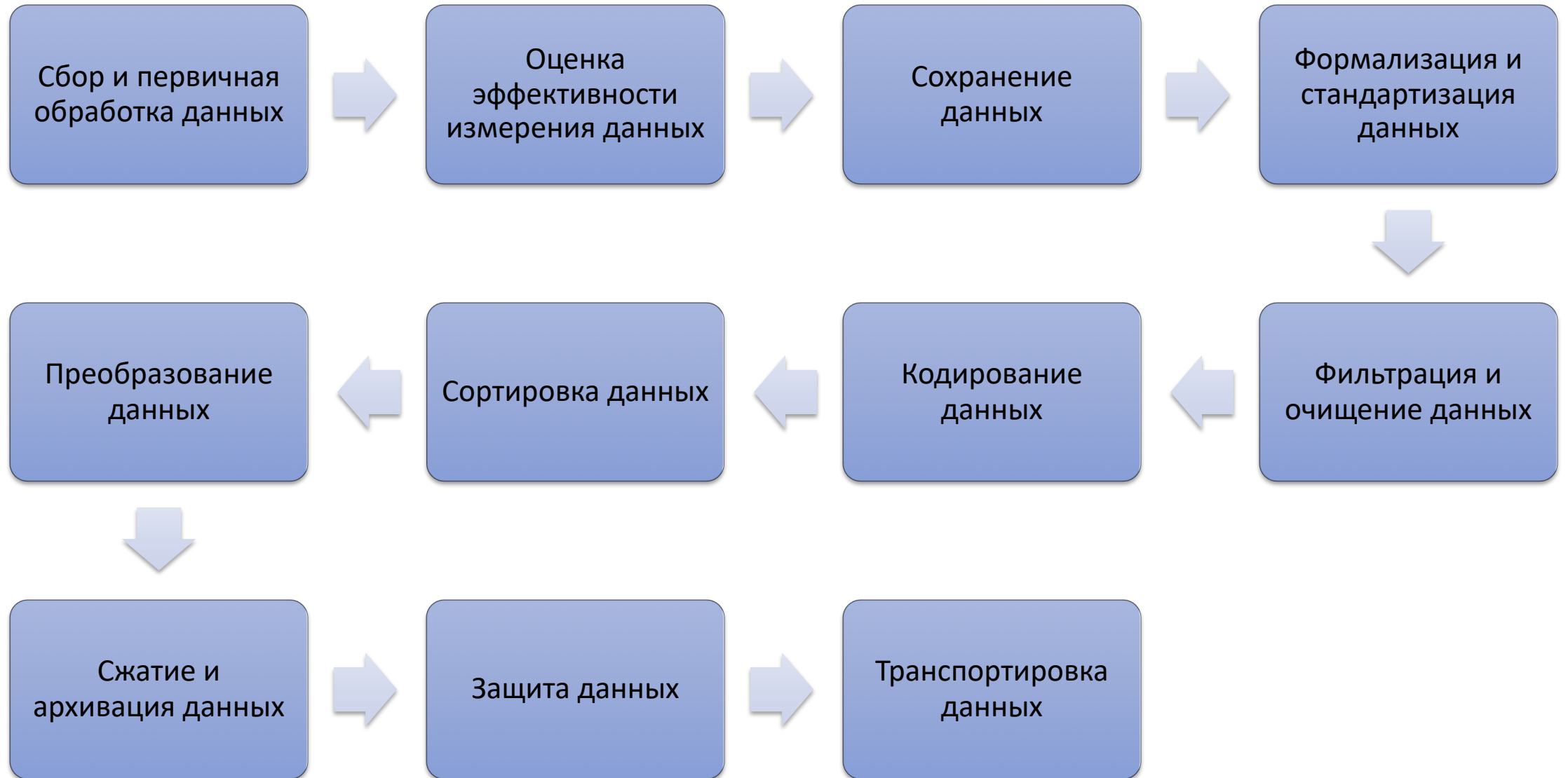
**Шкала наименований** – это группировка объектов и их производных в ряд непересекающихся классов (симптомы заболеваний и синдромы заболеваний).  
Например: классификация патологических затемнений легочных полей на рентгенограмме грудной клетки)

**Шкала порядка** – это упорядоченная шкала наименований, на которой отражена, в основном, тенденция процесса (например, степень желтушности кожных покровов, скорость оседания эритроцитов крови)

**Интервальная шкала** – это шкала с наличием единицы измерения (например, шкала температуры термометра)

**Шкала отношений** – это интервальная шкала с нулевой точкой (например, ростомер, весы)

# Этапы работы с медико-биологическими данными



# Этапы работы с медико-биологическими данными



# Действующие международные стандарты OSI-ISO\*

- **SCP-ECG** – стандарт для обмена цифровыми ЭКГ внутри и между лечебными учреждениями
- **DICOM 3.0** – стандарт обмена медицинскими изображениями (рентгенограмма, ультразвуковые и магнитно-резонансные томограммы).
- **IHE** – стандарт интеграции информационных систем.
- **HL7 (FAQ, News)** – стандарт обмена медицинскими данными.
- **ASCI X12** – стандарт обмена электронными документами.
- **IEEE P1157 («MEDIX»)** – стандарт обмена медицинскими
- данными.
- **CDA** – стандарт архитектуры клинических документов.
- **ASTM E3.11** – стандарт обмена данными лабораторных тестов.
- **CCOW** – стандарт клинического контекста.

*\*OSI (Open System Interconnection) – стандартизация открытых систем*

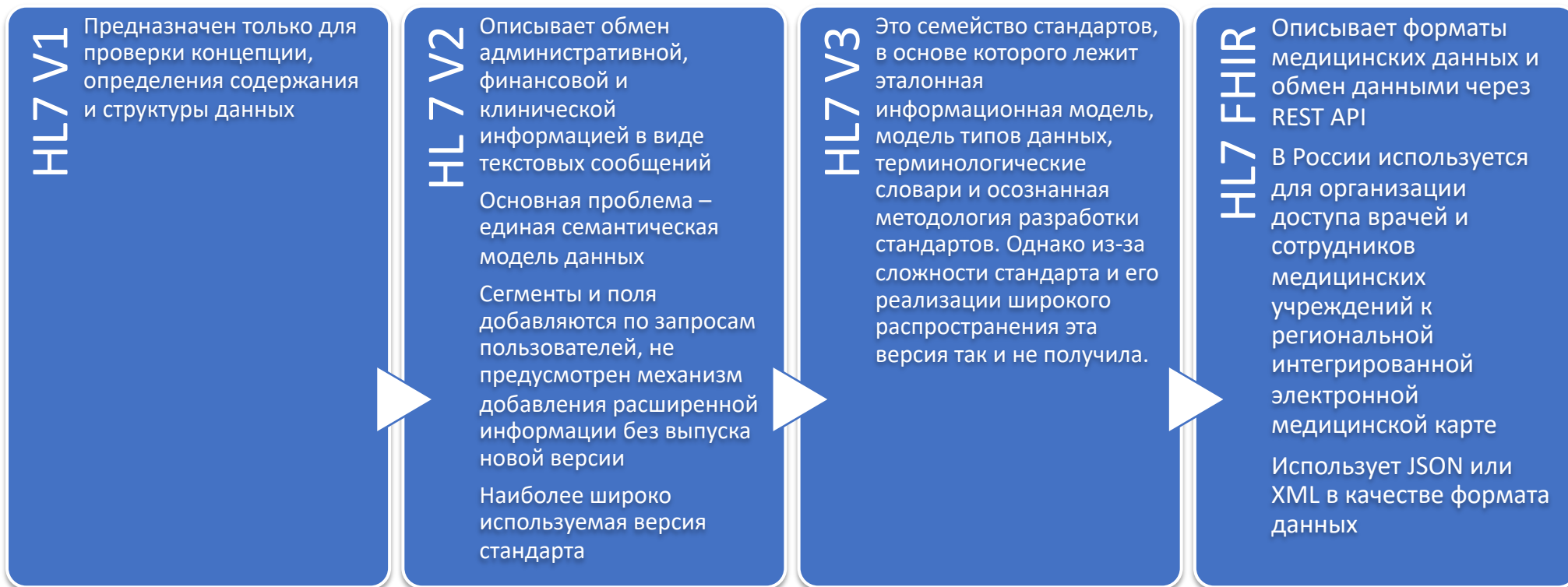
*ISO (International Standards) – Международный комитет по стандартизации*

# Стандарт HL7 (Health Level Seven)

HL7 – стандарт обмена, управления и интеграции электронной медицинской информации.

Поддерживает выполнение следующих задач:

- ✓ Структурирование передаваемых данных
- ✓ Возможности проектирования систем
- ✓ Достижение согласованности передач
- ✓ Безопасность
- ✓ Идентификация участников
- ✓ Доступность

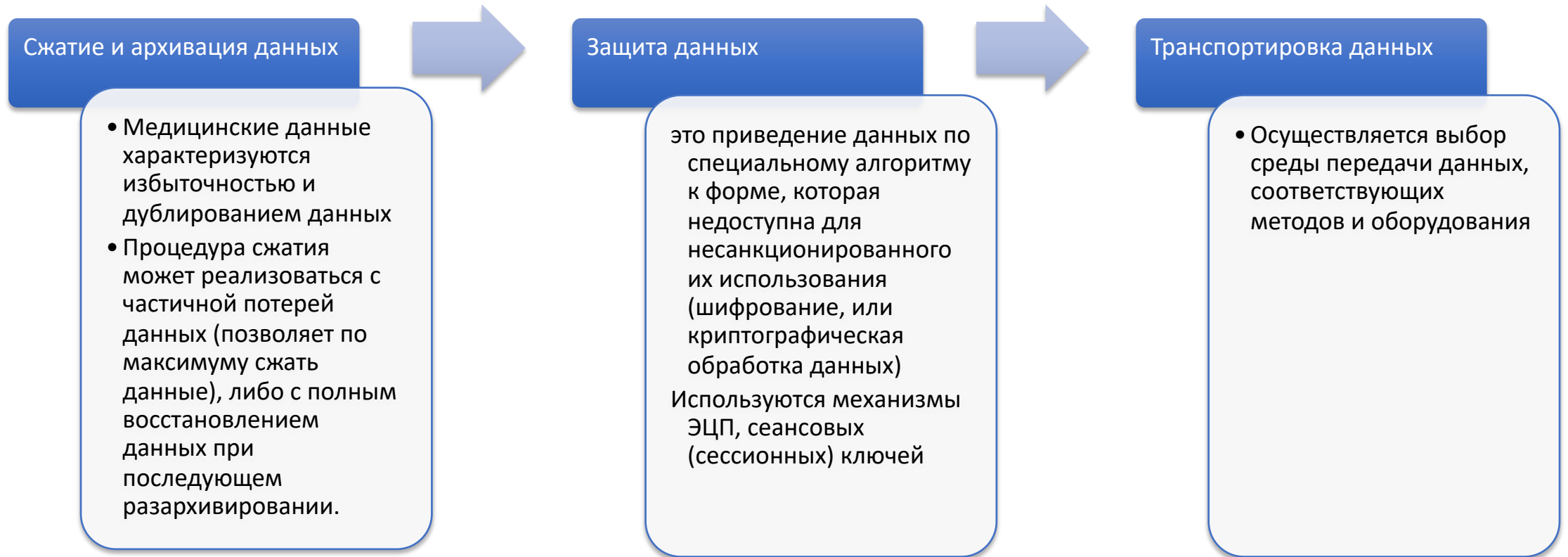




# Этапы работы с медико-биологическими данными (продолжение)



# Этапы работы с медико-биологическими данными (продолжение)



# Модели данных для хранения медицинской информации

**Структурированные данные** – данные, которые могут быть описаны некоторой моделью, определяющей их обработку (пример, реляционные модели данных, хранящие количественные данные в виде таблиц)

**Неструктурированные данные** – данные, не соответствующие ни одной модели, либо ее невозможно формализовать в явном виде (пример, нереляционные базы данных NoSQL, хранящие качественную информацию)

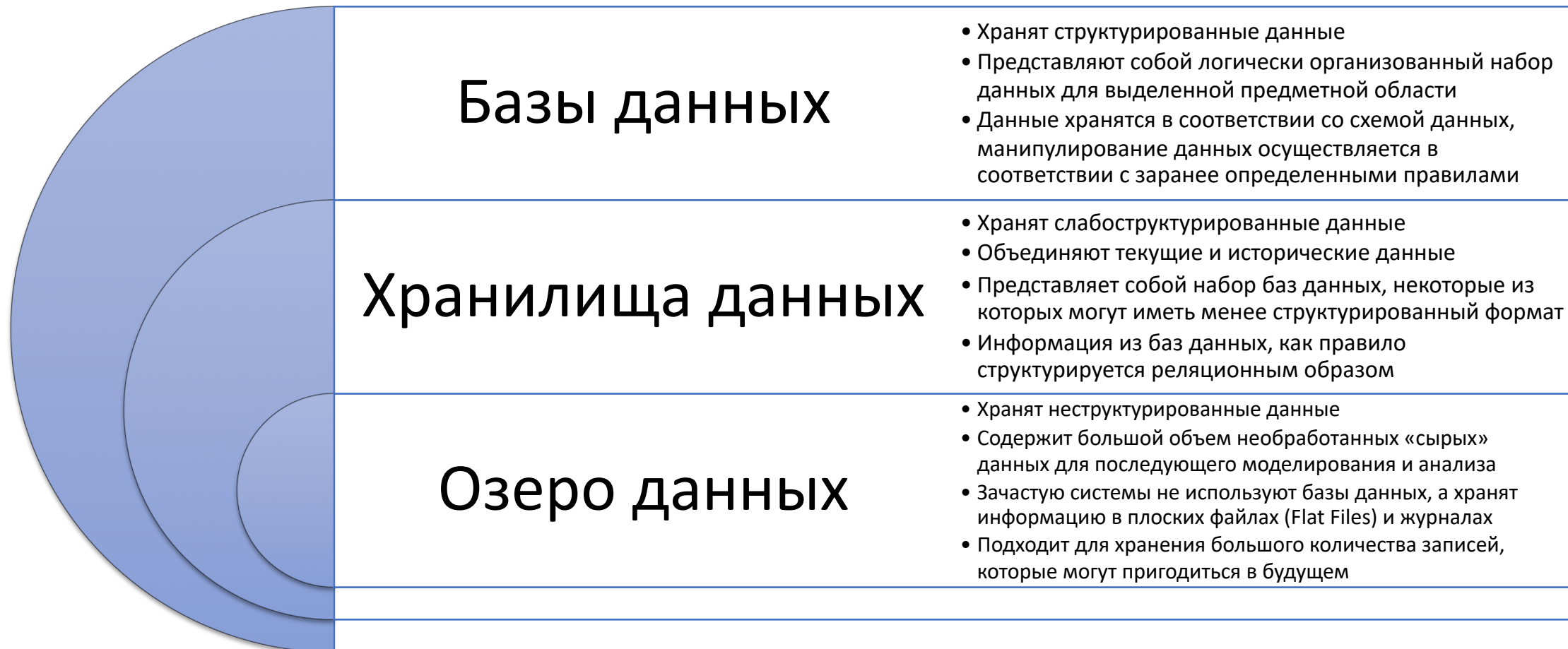
**Слабоструктурированные данные** – это форма структурированных данных, не соответствующая строгой структуре таблиц и отношений в моделях реляционных баз данных, но содержащая иные маркеры для определения семантических элементов и для обеспечения иерархической структуры записей и полей в наборе данных (пример, XML, html, JSON)

**Пример слабоструктурированных данных в медицине!** Запись в медицинской карте: для одного пациента имеется перечень прививок, для другого – показатели роста и веса, для третьего – операции, которые ему сделали.

# Особенности слабоструктурированных данных:

1. Отсутствует фиксированная схема данных
2. Отсутствует четкое различие между данными и их схемой
3. Отсутствует строгая типизация данных
4. Изменение схемы данных представляет собой рутинную операцию, сравнимую с внесением изменений в данные
5. Объем данных сравним со сложностью схемы
6. Схема данных является описывающей, а не предписывающей и может быть получена из самих данных
7. Полное знание схемы данных не является необходимым для построения запросов, возможны запросы, полностью игнорирующие схему данных

# Методы хранения данных в МИС



# Подходы к извлечению информации из слабоструктурированных данных

- Распознавание именованных сущностей (NER-Named Entity Recognition). Данные методы основаны на правилах и статистических методах анализа. Методы на основе правил требуют создания множества правил. Методы на основе статистики осуществляется на основ анализа рядом стоящих объектов
- Извлечение отношений (RE-Relation Extraction) – в основе лежит задача обнаружения семантических отношений между сущностями в тексте и характеристик. Поиск решений осуществляется через задачи классификации сущностей, либо определение степени схожести между объектами

# Слабоструктурированные данные – это большие данные!

К ним применимы технологии управления и анализа больших объемов данных.

Для сбора и анализа больших данных используются технологии **Data Mining**

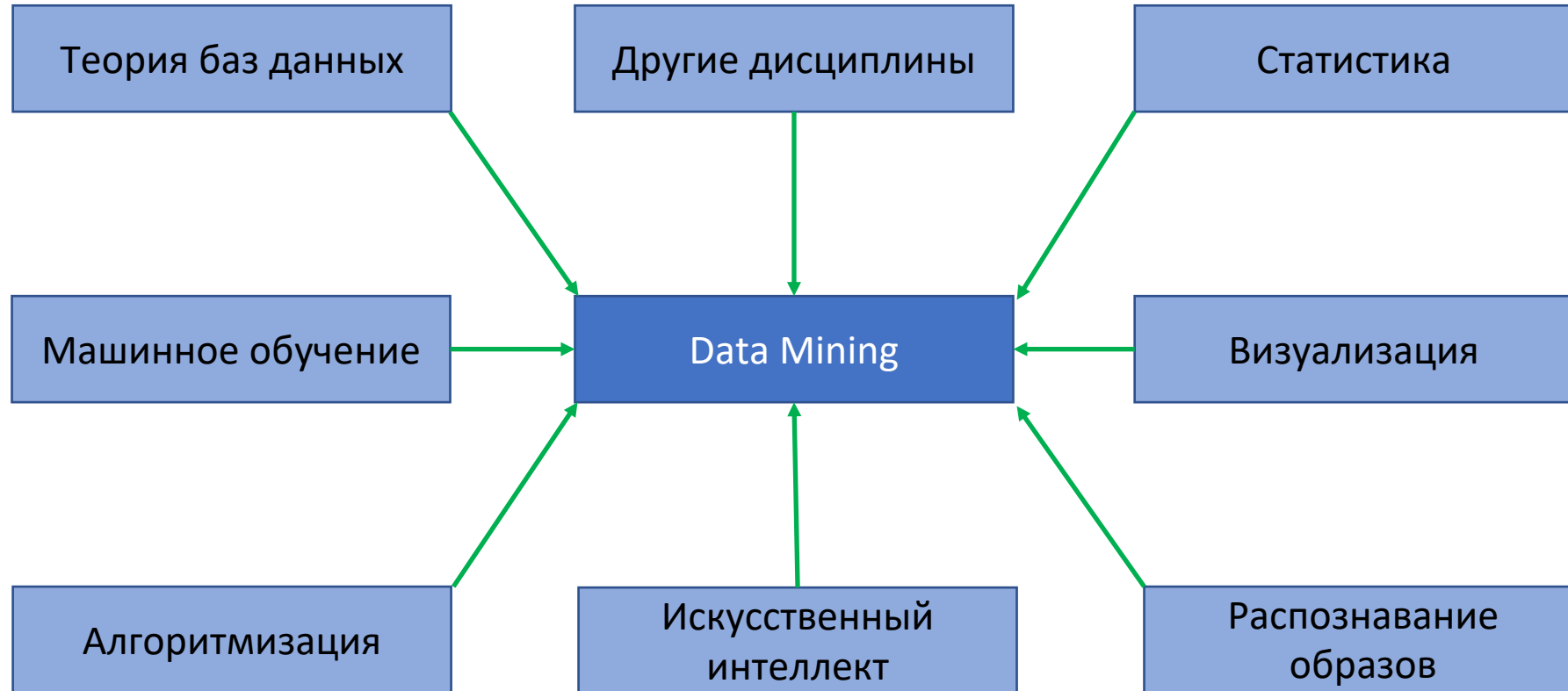
**Data Mining (добыча, интеллектуальный анализ данных, глубинный анализ данных)** – совокупность методов обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности:

**Неочевидных** – найденные закономерности не обнаруживаются стандартными методами обработки информации

**Объективных** – обнаруженные закономерности будут полностью соответствовать действительности, в отличие от экспертного мнения, которое всегда является субъективным

**Практически полезных** – выводы имеют конкретное значение, которому можно найти практическое применение

# Основа Data Mining





# Принцип работы Data Mining



Спасибо за внимание!