



Ассоциация  
Международных  
Фармацевтических  
Производителей

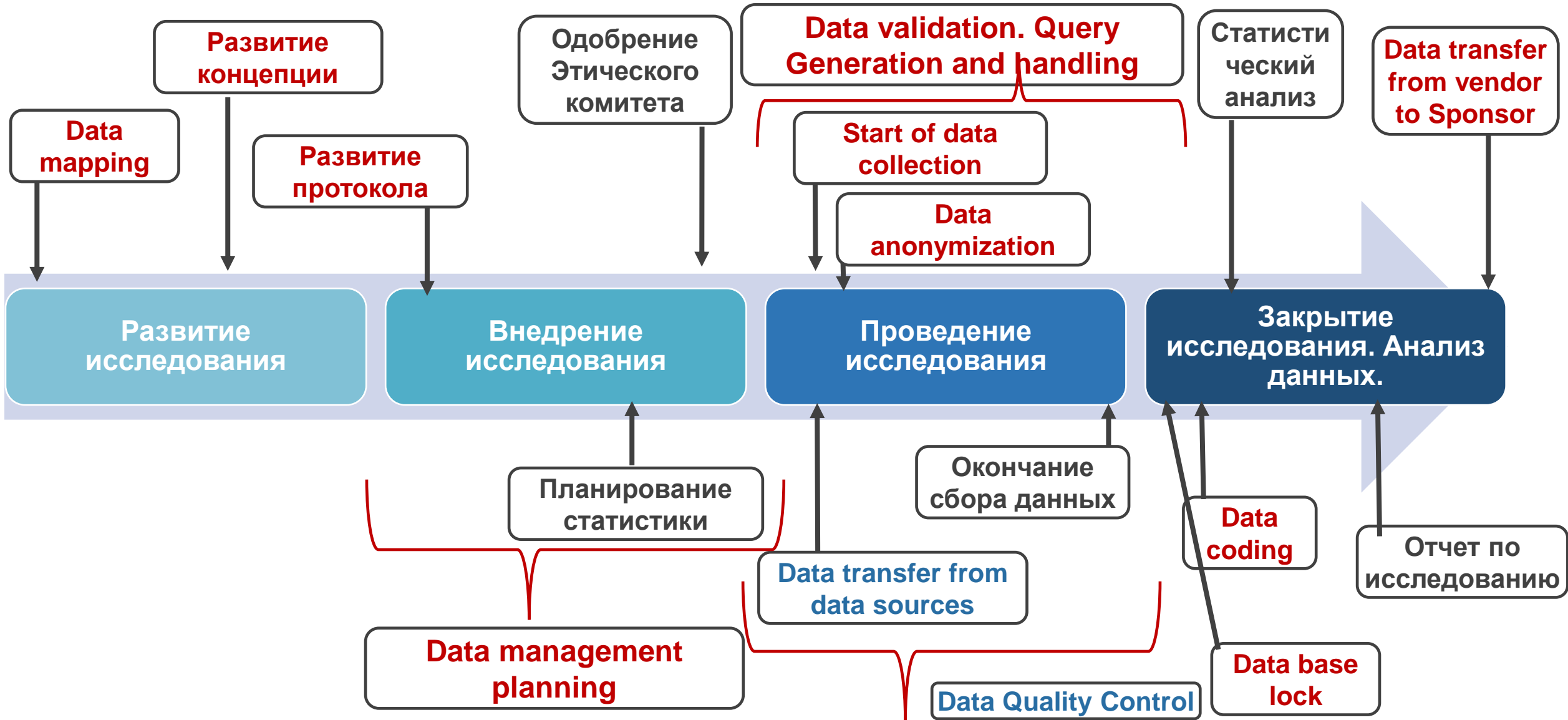
# Real World Evidence: управление данными

Проект «Диалоги о RWD/RWE»

Гольдина Татьяна, к.б.н.,  
руководитель рабочей группы AIPM по RWE

10 февраля 2023г

# Процессы управления данными на протяжении исследования



## Классификация исследований рутинной практики в зависимости от используемых критериев

Критерий	Типы и подтипы	Ссылки
Гипотеза/цель исследования	Поисковые и проверяющие; описательные и аналитические	[14, 32]
По наличию интервенции	Интервенционные (экспериментальные): прагматические РКИ; исследования в рамках когорты (trial within cohort (TwiCs) или cohort multiple RCT)*; всенаправленные когортные исследования (comprehensive cohort study (CCS)); РКИ с обогащением популяции (Population enrichment RCT); кластерные РКИ (Cluster RCT); нерандомизированные контролируемые исследования; большие упрощенные РКИ; гибридные исследования с внешней группой контроля (single-arm clinical trials with external controls). Квазиэкспериментальные: анализ прерванных временных рядов (interrupted time-series analysis); метод сравнения разностей (difference-in-difference); анализ до/пре- и после/пост (before-and-after). Неинтервенционные (наблюдательные): когортные; исследования случай—контроль; кросс-секционные исследования; исследования серии случаев с контролем по самим себе (self-controlled case series design); перекрестные исследования случай—контроль; описания клинических случаев	[23, 24, 30, 32]
Группа контроля/сравнения	Неконтролируемые; контролируемые: с внутренней группой контроля; с внешней группой контроля; с синтетической группой контроля исследования с контролем по самим себе; перекрестные исследования случай—контроль	[24, 32]
Рандомизация	Рандомизированные; квазирандомизированные; нерандомизированные	[24]
По сбору/типу используемых данных	Первичные данные; вторичное использование данных; гибридный дизайн	[*, 23, 30, 33]
Кратность сбора данных/по времени наблюдения участников/по времени прослеживания причинно-следственной связи	Кросс-секционные/одномоментные/поперечные; динамические/продольные/лонгитюдные; проспективные; ретроспективные; двунаправленные (ambidirectional, про-и ретроспективные)	[14, 23, 24, 30, 33, 34]

Новодережкина Е.А., Гольдина Т.А. Исследования рутинной практики: проблемы терминологии и классификации.

*Медицинские технологии. Оценка и выбор.*

2023;45(1):50–62.

<https://doi.org/10.17116/medtech20234501150>

## «ДО»

клинические исследования

неинтервенционные  
исследования  
(проспективные и  
ретроспективные)

- Рост числа источников данных
- Рост цифровых технологий



## «ПОСЛЕ»

клинические исследования

Исследования  
рутинной  
практики

Клинические  
исследования

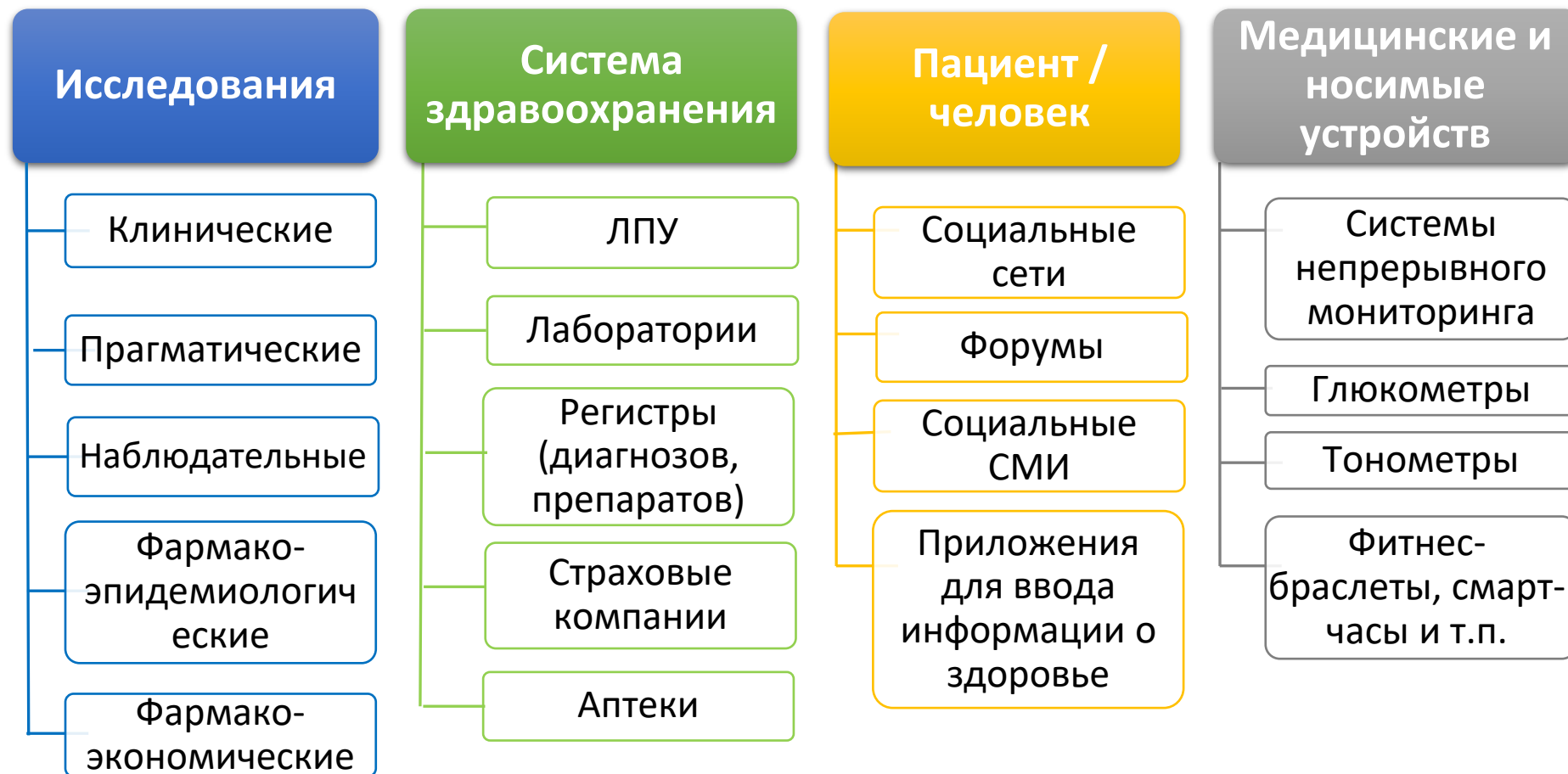
Неинтервенционные  
исследования

Иные варианты  
дизайна

Новодережкина Е.А., Гольдина Т.А. Исследования рутинной практики: проблемы терминологии и классификации. *Медицинские технологии. Оценка и выбор*. 2023;45(1):50–62. <https://doi.org/10.17116/medtech20234501150>

[Исследования рутинной практики: проблемы терминологии и классификации \(mediasphera.ru\)](https://mediasphera.ru)

# Источники RWD (Data sources)



## 2 ключевых пути работы с источниками данных

### Investigator study

Исследователь  
переносит данные  
вручную из  
источника в базу  
данных  
исследования

Разработка  
Индивидуальной  
Регистрационной  
Карты (ИРК)

Перенос  
данных  
вручную

Валидация  
данных

### Data base study

Прямой трансфер  
из источников  
данных

Формирование  
списка данных

Валидация  
данных

# Системы сбора данных в зависимости от дизайна RWE-исследований

по принципу интервенции

интервенционные

гибридный дизайн



Разработка  
ИРК

Формирование  
списка данных

у КИО часто есть  
шаблоны форм  
ИРК

неинтервенционные

первичные данные

вторичное  
использование данных

проспективные

ретроспективные

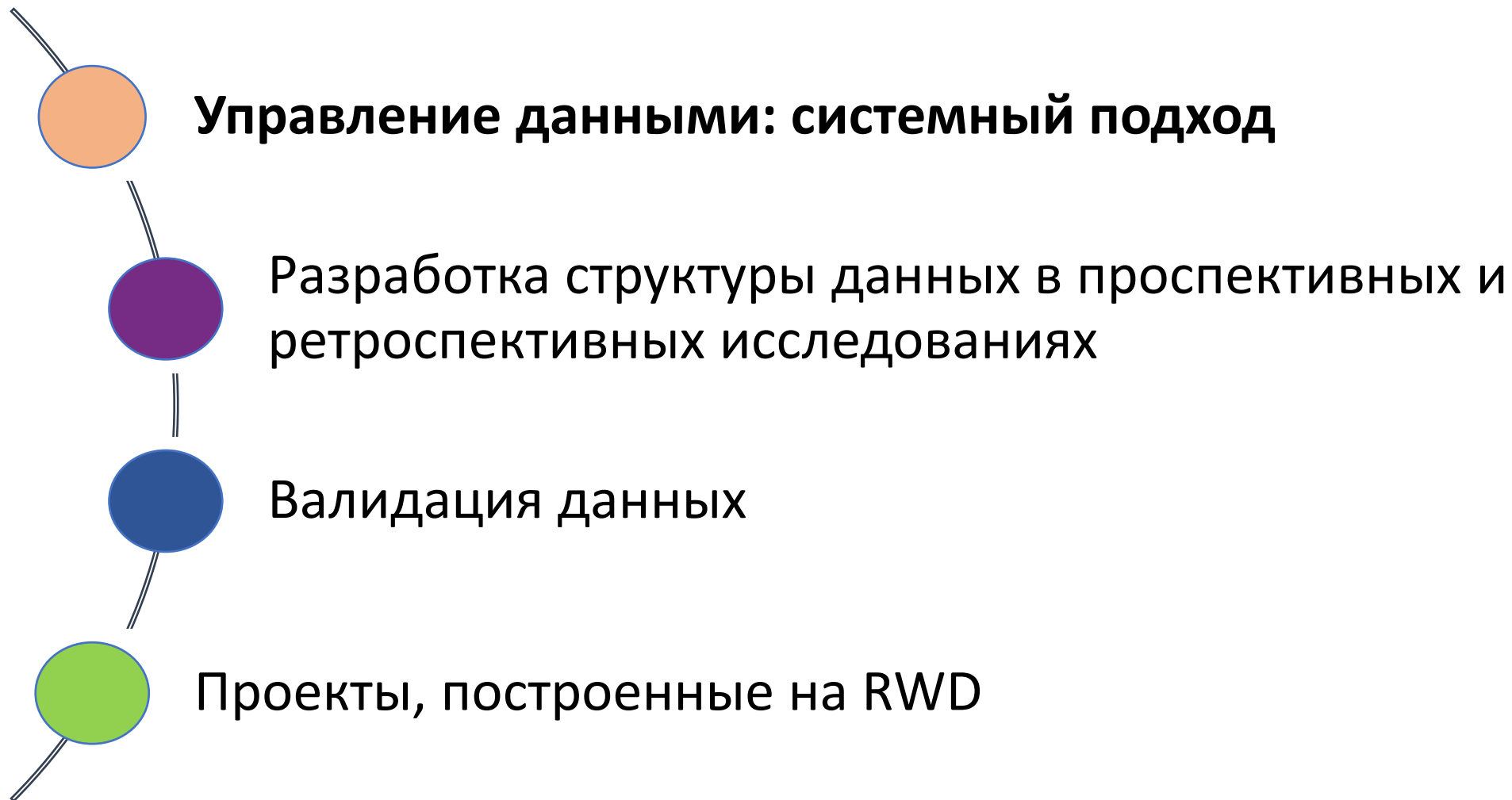


Разработка  
ИРК

Разработка  
ИРК

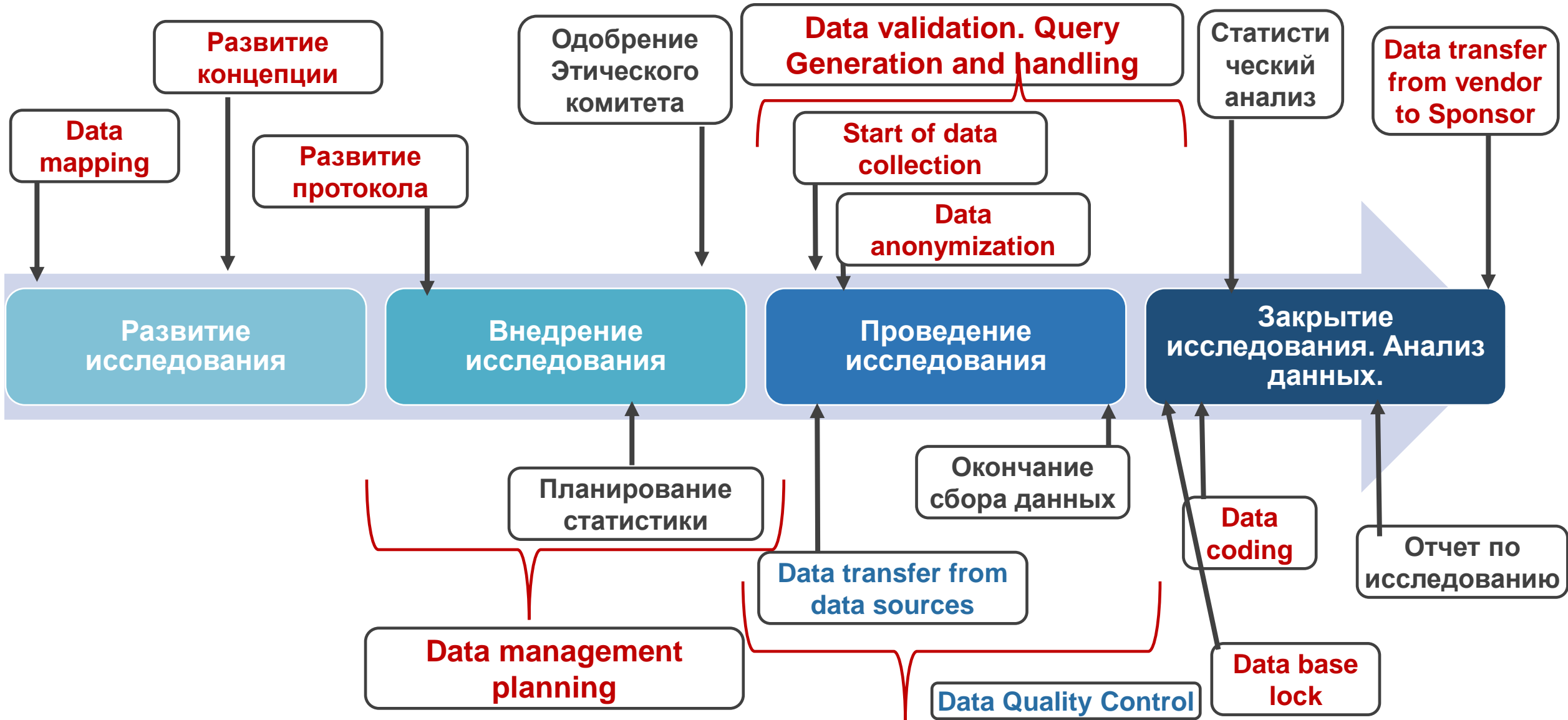
Формирован  
ие списка  
данных

Chart-review  
study





# Процессы управления данными на протяжении исследования

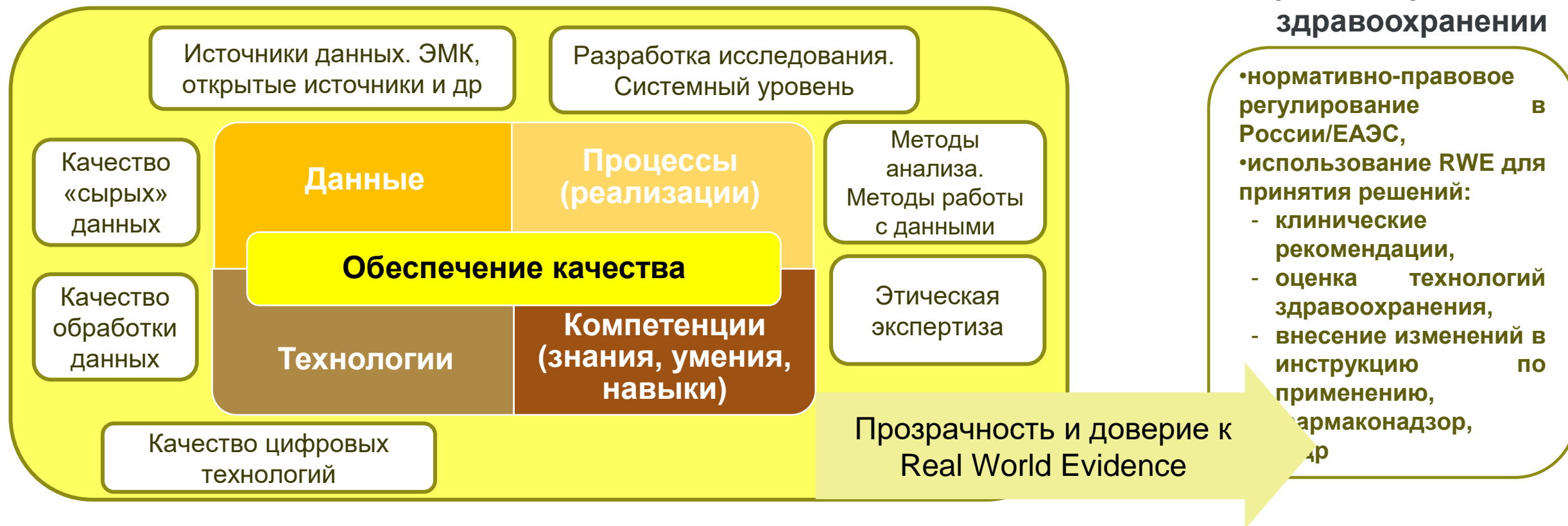


## RWD studies/RWD based projects

## RWD projects outcomes



## Принятие решений в здравоохранении



# Процессы управления данными

## Основные

- В процессе преобладает управление данными

## Составляющие

- Управление данными является составляющей частью других процессов

Развитие  
исследования

Data  
mapping

Развитие  
концепции

Внедрение  
исследования

Планирование  
управления  
данными

Разработка  
ИРК

Развитие  
протокола

Проведение  
исследования

Валидация  
данных

Управление отклонениями  
от протокола

## Управление данными

### Документы СМК Спонсора

- стандарты
- СОПы

+

### Документы СМК вендора(ов)

- стандарты
- СОПы
- рабочие инструкции

+

### Планы исследования

- Data Management Plan
- Data Validation Plan
- Data Transfer Plan
- И др

Описание всех требований и процессов для исследования

- **Data Management Plan (План управления данными)**
- **Data Validation Plan (План валидации данных)**
- **Data Transfer Plan (План передачи данных)**
- **Target Review Plan (План целевой проверки данных)**
  
- Protocol Deviation Management Plan (План управления отклонениями от протокола)
- Safety Management Plan (План управления данными по безопасности)
- Clinical Monitoring Plan (План мониторинга)
- Medical Monitoring Plan (План медицинского мониторинга)
- Quality Assurance Plan (План обеспечения качества) / Quality Control Plan (План контроля качества)
- Statistical Analysis Plan (План статистического анализа)
- etc

<b>Group of Processes/Process</b>	<b>Документы СМК Спонсора</b>	<b>Документы СМК Вендора-1</b>	<b>Документы СМК Вендора-2</b>
<b>General requirements</b>			
<b>Study documentation</b>			
<b>Ethical approval</b>			
<b>Study management</b>			
<b>Data management</b>			
<b>Trial Master File</b>			



## Синописис

- Определение потребностей
- Целевая популяция
- Цели и конечные точки
- Критерии включения/невключения
- Цели и конечные точки
- Основные собираемые данные
- План сбора данных

## Протокол

- Критерии включения/невключения
- Цели и конечные точки
- План наблюдения
- Сбор данных (перечисление всех собираемых данных)

## ИРК

## Data mapping

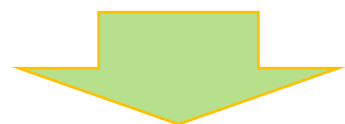
to estimate do we have data to realize a study



# RWY study objectives, end points and RWD – useful tool

Level	Study Objective	Study End points	Data needed to receive evidence			
			Baseline	Month 3	Month 6	Data source
	To evaluate HbA1c level dynamic at month 6 in total population	Percentage of patients achieved individual target HbA1c defined by treating physician by Month 6	predefined individualized HbA1c level	N/A	HbA1c	
		Change From Baseline in Glycosylated Hemoglobin (HbA1c) Level at Month 6	HbA1c	N/A	HbA1c	
	To evaluate HbA1c level dynamic over Time at month 6 in total population	Change From Baseline in HbA1c Level Over Time	HbA1c	HbA1c	HbA1c	

The key task is to match data and end points



As a result: listings of required data

Data needed to receive evidence			
Baseline	Month 3	Month 6	Data source
predefined individualized HbA1c level	N/A	HbA1c	predefined - hospital, HbA1C-VLM
predefined individualized HbA1c level	HbA1c	N/A	paper medical chart
predefined individualized HbA1c level and HbA1c	HbA1c	HbA1c	paper medical chart



## Data collection overview

	Index date	Month 3	Month 6
Data			
Targeted HbA1c, FPG and PPG	+		
HbA1c, FPG and PPG	+	+	+

## 3 варианта:

- Бумажная, БИРК (paper, pCRF)
- Электронная, эИРК (electronic. eCRF)
- Гибридная (hybrid)

## Электронная система сбора данных (EDC):

### Внесение данных

- Исследователем - в электронную ИРК (eCRF)
- Пациентом - в электронный дневник, опросник (ePRO)

**Медицинский анамнез**  
Имеются ли данные о сопутствующих заболеваниях за 180 дней до наступления индексного события?

Укажите: I\_SOL02\_MHYN \* Если "Да", заполните форму "Сопутствующие заболевания" в разделе "Приложения"

(выберите значение)  
Да [Y]  
Нет [N]  
Нет данных [NA]

---

**Анамнез СД2**  
Дата первичного (первого за все время) диагноза диабета I\_SOL02\_T2D\_MHSTDAT \*

Имя:

---

**Осложнения диабета**

Нейропатия I_SOL02_NEU_MHOCCUR *	Дата постановки диагноза I_SOL02_NEU_MHSTDAT <input type="text"/> <input type="calendar"/>	Продолжается на момент наступления индексного события? I_SOL02_NEU_MHONGO (выберите значение) Да [Y] [Y] Нет [N] [N] Нет данных [NA] [NA]
Нефропатия I_SOL02_NEP_MHOCCUR *	Дата постановки диагноза I_SOL02_NEP_MHSTDAT <input type="text"/> <input type="calendar"/>	Продолжается на момент наступления индексного события? I_SOL02_NEP_MHONGO (выберите значение) Да [Y] Нет [N] Нет данных [NA]
Ретинопатия I_SOL02_RET_MHOCCUR *	Дата постановки диагноза I_SOL02_RET_MHSTDAT <input type="text"/> <input type="calendar"/>	Продолжается на момент наступления индексного события? I_SOL02_RET_MHONGO (выберите значение) Да [Y] Нет [N] Нет данных [NA]

Собираем те данные, которые необходимы для достижения результатов (evidence) исследования

- Конечные точки
- Субанализы
- Объяснение, обсуждение результатов

Учитываем особенности RWD

Максимальная экономия ресурсов исследователя

Применение правил социологических исследований

- Ясные вопросы, дающие однозначное понимание
- Простые формулировки

## НЕТ ВМЕШАТЕЛЬСТВА – ОЦЕНКА РУТИННОЙ КЛИНИЧЕСКОЙ ПРАКТИКИ

### Ключевой вопрос: как ... в рутинной клинической практике?

Например:

- Что описано в руководстве по лечению заболевания и как это происходит в рутинной клинической практике?
- Какие типы/паттерны лечения я знаю в рутинной клинической практике?
- Какие пациенты в рутинной клинической практике?
- Как часто пациенты в рутинной клинической практике ходят на визиты к врачу?/Как долго в рутинной клинической практике длится госпитализация
- И т.д.

## 2 типа форм ИРК

### Единоразовая форма сбора

Собирает данные в конкретный момент времени исследования



- Формы визитов (Визит включения, Визит 3 месяца от включения и тд)
- Анамнез заболевания
- Окончание сбора данных
- И тп.

### Повторяющаяся форма сбора

Собирает данные в течение определенного времени проведения исследования



- Лабораторные анализы
- Сопутствующая терапия
- Сопутствующие заболевания

Соотношение и состав полей ИРК определяется в каждом исследовании индивидуально, в зависимости от научных задач и имеющихся данных

Форма

F\_LAN01\_3313

Секция

Включен... (0/14)

Заголовок:

Инструкция:

Начало визита

Дата визита I\_LAN01\_VISDAT\_2050 \*



Демографические данные

Дата рождения I\_LAN01\_BRTHDAT\_4111 \*



Возраст, полных лет I\_LAN01\_AGE\_5448

автоматический расчет возраста

Пол I\_LAN01\_SEX\_5352 \*

(выберите значение)

Мужской [M]

Женский [F]

Антропометрические данные

Вес, кг I\_LAN01\_WEIGHT\_VSORRES\_5425 \* Результат округляется до десятых долей килограмма

Рост, см I\_LAN01\_HEIGHT\_VSORRES\_910 \* Результат округляется до ближайшего сантиметра

ИМТ I\_LAN01\_BMI\_VSORRES\_4719 автоматический расчет ИМТ





# 4 способа ввода данных (1)

Поле, свободное для записи

Дата

Возраст на момент наступления индексного события I\_SOL02\_AGE (лет)

Препарат  
 Код

Форма LAN80 Осложнения СД  
 Секция Терапия осложнений СД

Осложне...(0/7)    Терапия...(0/14)    -- Выбрать для переход

Заголовок: Терапия осложнений СД

Удалено	Торговое название препарата	
I_LAN80_INVALIDATED_01 <input type="checkbox"/> удалено	I_LAN80_CMTRT	I_LAN80_CMINGR

Добавить

Диализ, дата начала осложнения: I\_LAN80\_DIALYS\_MHSTDAT



Ампутация, дата начала осложнения: I\_LAN80\_AMP\_MHSTDAT



Язва стопы, дата начала осложнения: I\_LAN80\_ULC\_MHSTDAT



## 4 способа ввода данных (2)

### Множественный выбор

**Микрососудистые осложнения**

Имеются ли данные I\_LAN80\_MICVASCOM\_MHYN \*  
о микрососудистых осложнениях сахарного диабета?

(введите значение)  
Да [Y]  
Нет [N]

Уточните I\_LAN80\_MICVASCOM\_MHTERM  
осложнение:

- Ретинопатия
- Сенсорная нейропатия (нарушение чувствительности)
- Микроальбуминурия
- Протеинурия
- Диализ
- Ампутация
- Язва стопы (активная или в анамнезе)
- хПБ
- Другое

### Единичный выбор (Выпадающий список)

Расовая I\_SOL02\_CRACE \*  
принадлежность

(выберите значение)

- Европеоидная [1]
- Негроидная [2]
- Монголоидная [3]
- Смешанная [указать] [4]
- Другая [указать] [5]
- Невозможно установить [6]

**Частота применения**

I\_LAN80\_DMCOM\_CMDOSFRQ

(введите значение)

- 1 раз в день [1]
- 2 раза в день [2]
- 3 раза в день [3]
- через день [4]
- другое [5] \*

**Путь введения**

I\_LAN80\_DMCOM\_CMROUTE I\_

(введите значение)

- перорально [1]
- внутримышечно [2]
- внутривенно [3]
- подкожно [4]
- другое [5] \*

# Плюсы и минусы различных способов ввода данных

Поле, свободное для записи

Минусы:

- ошибки ввода,
- затрачиваемое время,
- «хаос» информации

Плюсы:

- удобство разработки,
- максимальный сбор информации

Множественный выбор

+

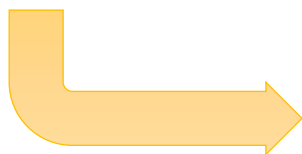
Единичный выбор

Минусы:

- пропустить параметр при разработке формы,
- листинг не более 7 позиций (особенности восприятия)

Плюсы:

- структуризация данных на этапе сбора данных,
- продумывание представления результата,
- упрощение заполнения исследователями



**Насколько это позволяет исследование и рутинная клиническая практика, выбираем форму «выпадающий список»**

## Сопутствующая терапия

Имеются ли данные о сопутствующей терапии за 180 дней до наступления индексного события?

Укажите: I\_SOL02\_CMYN

\* Если "Да", заполните форму "Сопутствующая терапия" в разделе "Приложения"

(введите значение)

Да [Y]  
Нет [N]  
Нет данных [NA]

## Получение медицинской помощи в связи с СД2

Имеются ли данные о получении медицинской помощи в связи с СД2 за 180 дней до наступления индексного события?

Укажите: I\_SOL02\_T2D\_HOYN

\* Если "Да", заполните форму "Получение медицинской помощи" в разделе "Приложения"

(введите значение)

Да [Y]  
Нет [N]  
Нет данных [NA]

## Эпизоды гипогликемии

Имеются ли данные об эпизодах гипогликемии за 180 дней до наступления индексного события?

Укажите: I\_SOL02\_HE\_CEOCCUR

\* Если "Да", заполните форму "Эпизоды гипогликемии" в разделе "Приложения"

(введите значение)

Да [Y]  
Нет [N]  
Нет данных [NA]

Пропущенные данные  
(missing data) – это  
норма в рутинной  
практике

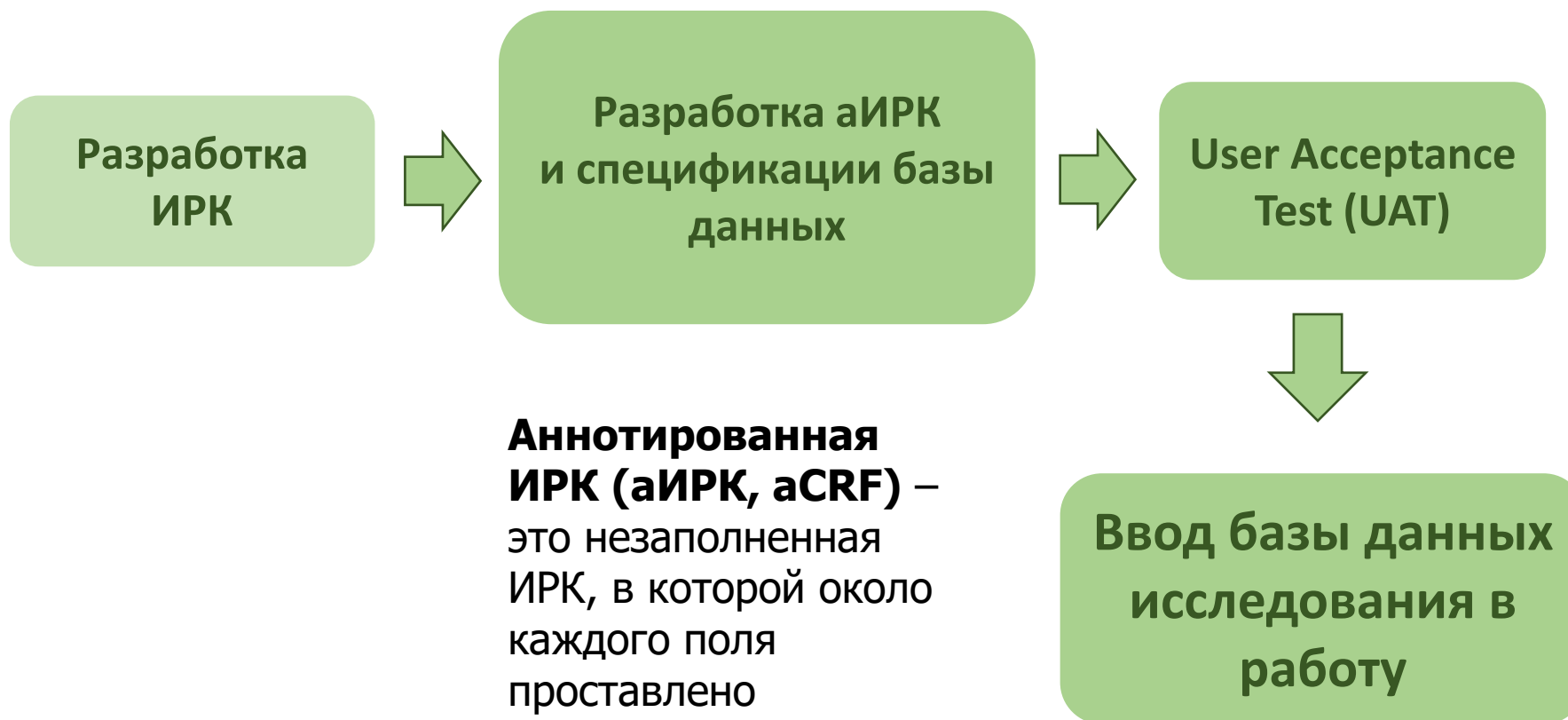


Если поле ИРК останется  
пустым (незаполненным), то  
непонятна причина (врач  
пропустил или данных нет)



Ввод в ИРК вопросов с  
ответом «да/нет/нет данных»

**Ошибки в разработке ИРК могут иметь  
очень серьезные последствия: потеря  
части (от небольшой до значимой)  
научных результатов**



**Аннотированная ИРК (аИРК, аCRF)** – это незаполненная ИРК, в которой около каждого поля проставлено название соответствующей переменной в базе данных.

## User Acceptance Test (UAT)

В базу данных вводятся вымышленные данные (test/dummy data) с целью тестирования:

- **Дизайна экранов eCRF:**
  - разделов
  - визитов
  - полей
- **Встроенных проверок (edit checks):**

Каждая проверка должна хотя бы раз пройти тестирование на случай:


  - Ошибки
  - Отсутствия ошибки



Commentary |  Open Access

## Guidelines for good database selection and use in pharmacoepidemiology research<sup>†</sup>




 Correction(s) for this article ▾

Gillian C. Hall  Brian Sauer, Alison Bourke, Jeffrey S. Brown, Matthew W. Reynolds, Robert Lo Casale

First published: 08 November 2011 | <https://doi.org/10.1002/pds.2229> | Citations: 95

<sup>†</sup> This document has been endorsed by the International Society for Pharmacoepidemiology as policy and has been posted on their website as a preprint version.

 SECTIONS

 PDF  TOOLS  SHARE

### ABSTRACT

The use of healthcare databases in research provides advantages such as increased speed, lower costs and limitation of some biases. However, database research has its own challenges as studies must be performed within the limitations of resources, which



References



Related



Information

### Recommended

[Use of Health Care Databases in Pharmacoepidemiology](#)

Sean Hennessy

Basic & Clinical Pharmacology & Toxicology

[A database selection expert system based on reference librarian's database selection strategy: A usability and empirical evaluation](#)

Wei Ma

Journal of the American Society for

1. Selection of a database,
2. Use of multiple data resources,
3. Extraction and analysis of the study population,
4. Privacy and security,
5. Quality and validation procedures and
6. Documentation

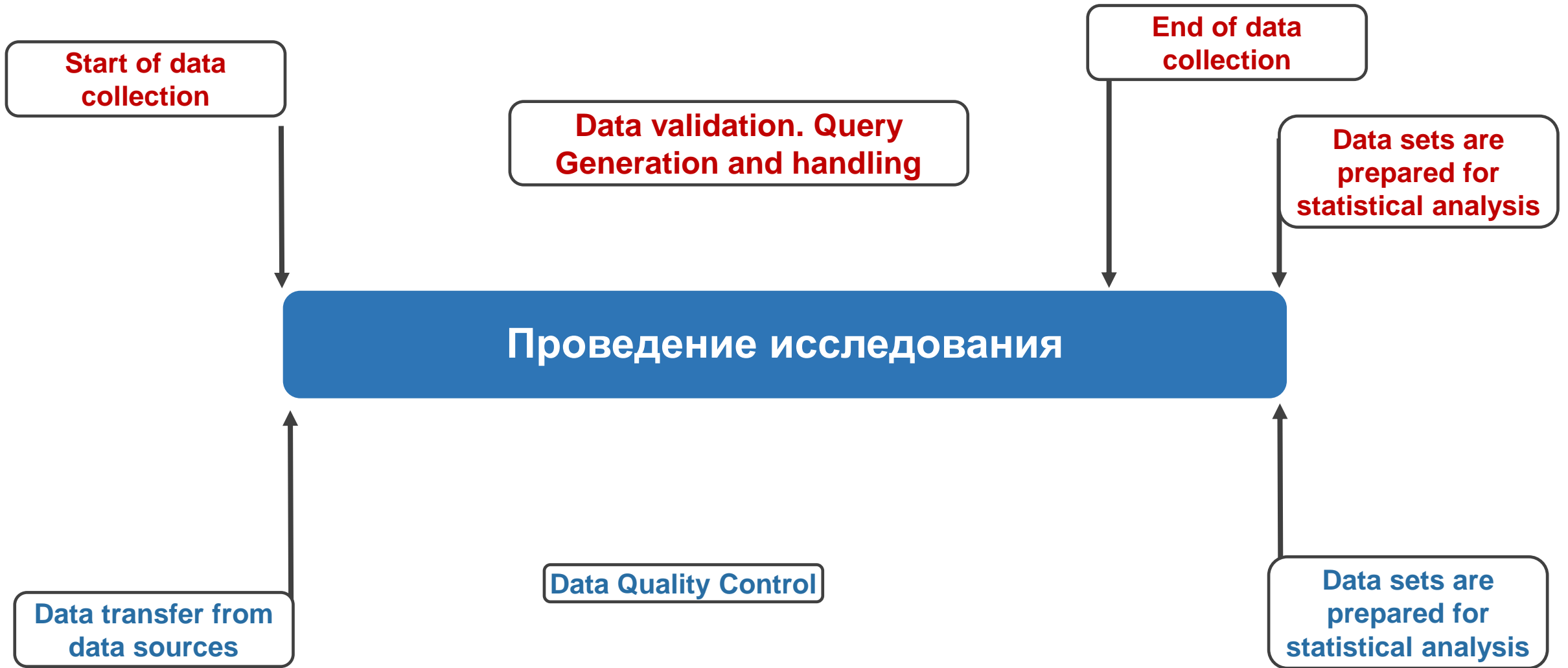
## Checklist for investigators in database research

<p style="text-align: center;"><b>3 Extraction and analysis of the study population</b></p> <p><b>Specification of extraction:</b> Are the following specified in detail: how to extract the study population and variables, code lists and non-coded systems, retrieval and merging of additional external data, output and final analysis?</p>
<p style="text-align: center;"><b>4 Privacy and security</b></p> <p><b>Compliance with privacy and security policy:</b> Have all relevant local, regional and national policies been complied with?</p> <p><b>Limited use of identifying information:</b> Are all direct identifiers removed or masked? Whose responsibility is it to ensure privacy?</p> <p><b>Secure data storage and transfer:</b> Is there a formal data security policy, and has this been adhered to?</p>





# Процессы управления данными на протяжении исследования



## План проверки достоверности данных:

- Подробное описание необходимых проверок для выявления неожиданных, противоречивых или несогласованных данных.
- Описание процесса обработки запросов
- Перечень ответственных лиц и соответствующих процессов (SOPs)
- Предоставляемая документация

## Проверки

- **Автоматические (Programmed edit checks):**
  - Стандартные - заложенные (DVS template checks) или программируемые в EDC системе
    - исправление ошибок на этапе вводе данных
  - Специфичные для конкретного протокола - программируемые в SAS (offline checks)
- **«Ручные» проверки (Manual checks):**
  - Сложные проверки
  - Проверки, требующие экспертной (субъективной) оценки.

Показатель	Расшифровка	OID
Height_VSORRES	Результат, см	I_SOL03_HEIGHT_VSORRES
Правило	E-mail	Визит
Сообщение	Ошибка валидационного правила: Рост указан меньше 140 см или больше 210 см, проверьте, пожалуйста	

Показатель	Расшифровка	OID
EXSTDAT	Дата начала терапии	I_SOL03_EXSTDAT
Правило	E-mail	Визит
Сообщение	Ошибка валидационного правила: Дата начала терапии препаратом X <sup>®</sup> указана раньше 01-11-2018 или позже 30-06-2020, проверьте критерии включения, пожалуйста	

Показатель	Расшифровка	OID
DSENDAT_DSDECOD	Дата окончания сбора данных	I_SOL05_DSENDAT_DSDECOD
Правило	E-mail	Визит
Сообщение	Ошибка валидационного правила: Дата окончания сбора данных указана раньше даты начала терапии или позже её более чем на 13 месяцев , проверьте, пожалуйста	

Показатель	Расшифровка	OID
OAD_CMDOSTOT	Суточная доза	I_SOL10_OAD_CMDOSTOT
Правило	E-mail	Визит
Сообщение	Ошибка валидационного правила: Указанная доза меньше или равна 0 либо больше 3000, проверьте, пожалуйста	

- Наличие сопутствующей терапии, а при этом не указаны сопутствующие заболевания;
- Наличие сопутствующих заболеваний, а при этом не указана сопутствующая терапия;
- Проверка всех показателей гликированного гемоглобина на согласованность данных сквозь время наблюдения
- Форма терапии СД – проверка согласованности сроков окончания терапии

## Проверки

- Автоматические (Programmed edit checks):
- «Ручные» проверки (Manual checks):

## Статистическая валидация

Показатель	Расшифровка	OID
I_LAN85_TAB_CMSTDAT	Дата начала	
Правило	E-mail	Визит
Сообщение	Дата начала таблетированной терапии СД указана позже даты окончания, проверьте, пожалуйста	
Выражение	PDATE(I_LAN85_TAB_CMSTDAT) > PDATE(I_LAN85_TAB_CMENDAT)	
Показатель	Расшифровка	OID
I_LAN86_INSSHORT_CMS TDAT	Дата начала	
Правило	E-mail	Визит
Сообщение	Дата начала инъекционной терапии СД указана позже даты окончания, проверьте, пожалуйста	
Выражение	PDATE(I_LAN86_INSSHORT_CMSTDAT) > PDATE(I_LAN86_INSSHORT_CMENDAT)	

# Quality control of data

Data source	Data from EDC	Data Base 1	Data from medical device
<b>Processes for Quality control</b>			
<b>Data validation and verification with data source</b>	yes (Data Validation Plan, section 2 )	Not applicable	Not applicable
<b>Statistical validation</b>	yes (Data Validation Plan, section 3.1. )	yes (Data Validation Plan, section 3.2. )	yes (Data Validation Plan, section 3.3. )
<b>Verification after statistical validation</b>	yes	Not applicable	Not applicable
<b>Statistical validation: cross-checks between data source</b>	yes (Data Validation Plan, section 4 )		
<b>Additional review of data via Microsoft R</b>	yes	yes	yes



## MedDRA® - для медицинских состояний:

- НЯ/СНЯ
- Явления до начала лечения
- Медицинская история
- Сопутствующие состояния;

## WHO-DD - для медикаментов:

- История приема препаратов (Pre-treatment)
- Сопутствующая терапия



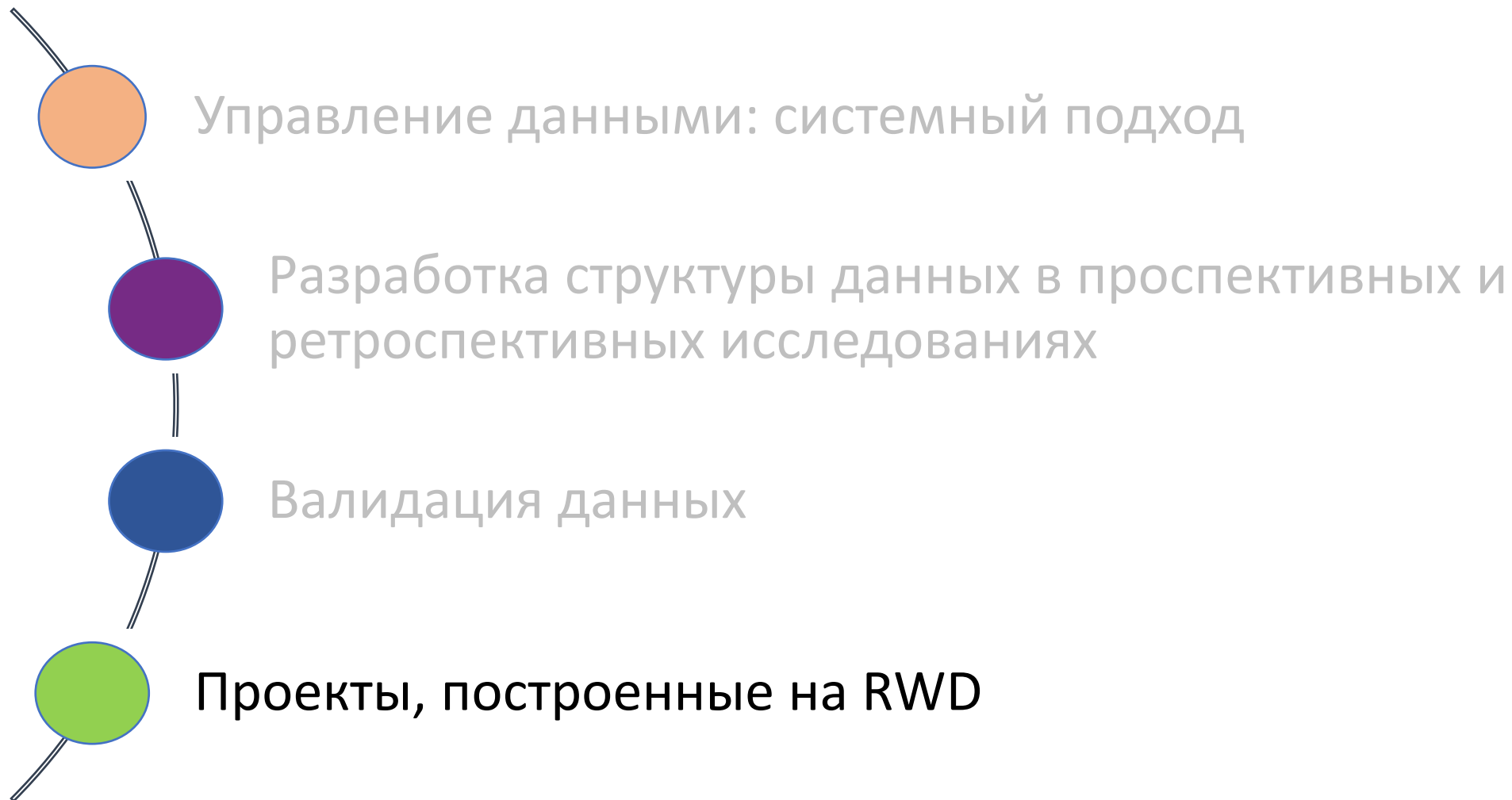
Анатомо-  
терапевтическо-  
химическая  
классификация (АТХ)

Кодирование осуществляется в соответствии с определенными правилами:

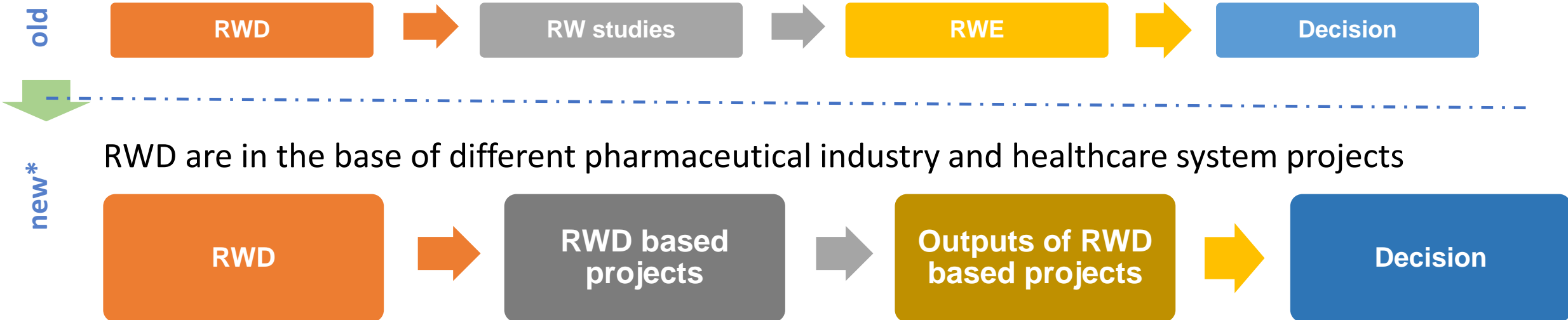
- Словарь MedDRA интегрируется в систему: кодирование полуавтоматическое (вручную выбирается LLT, остальные термины PT name, PT code, SOC name, SOC code система кодирует автоматически).
- Словарь WHO Drug Dictionary может интегрироваться в систему, но кодирование терминов, как правило, делается вручную (Class code, Class name and Group name).
- Если кодирование по словарям невозможно, высылается запрос исследователям
- Если кодирование по словарям невозможно, то вместо терминов указывается «Код не определен»
- Контроль кодирования осуществляется квалифицированным медицинским специалистом
- Со стороны Спонсора кодирование одобряется специалистами по фармаконадзору

# Заккрытие базы данных

- Заккрытие базы данных – этап, после которого никакие дальнейшие манипуляции с данными совершаться не могут. Т.е. можно проводить анализ данных.
- Заккрытие базы данных производится после того, как все этапы управления данными завершены, а именно:
  - Все данные введены, все формы ИРК подписаны исследователями;
  - Все формы, требующие сверки первичных данных, проверены специалистом по управлению данными и переведены в статус сверки первичных данных;
  - Валидация данных произведена в соответствии с Планом валидации данных,
  - Нет открытых запросов на уточнение данных; все проверки и запросы на уточнение данных разрешены;
  - Проведен контроль качества;
  - Кодирование данных произведено
- Передача базы данных для анализа

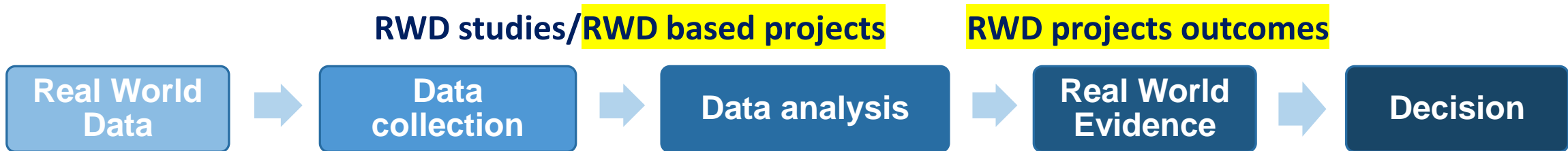


# Paradigm change: from RWE to RWD based projects



RWD are in the base of different pharmaceutical industry and healthcare system projects

- Electronic Health Records (incl treatment, health care utilization data)
- Patient data from open sources
- Medical devices data
- Internal Sanofi data (studies, programs)
- External data lakes
- Real World Evidence (RWE)
- Clinical decision support system (CDSS)
- Patient Screening
- Diagnostic projects
- Market Research
- Different digital projects
- etc



***Спасибо за внимание!***