

Business Understanding. С чего начинается  
работа с данными?

Skillbox

# Методология CRISP-DM

“

Сегодня в голове,  
завтра в чертежах,  
послезавтра в металле,  
а потом и в полёте.

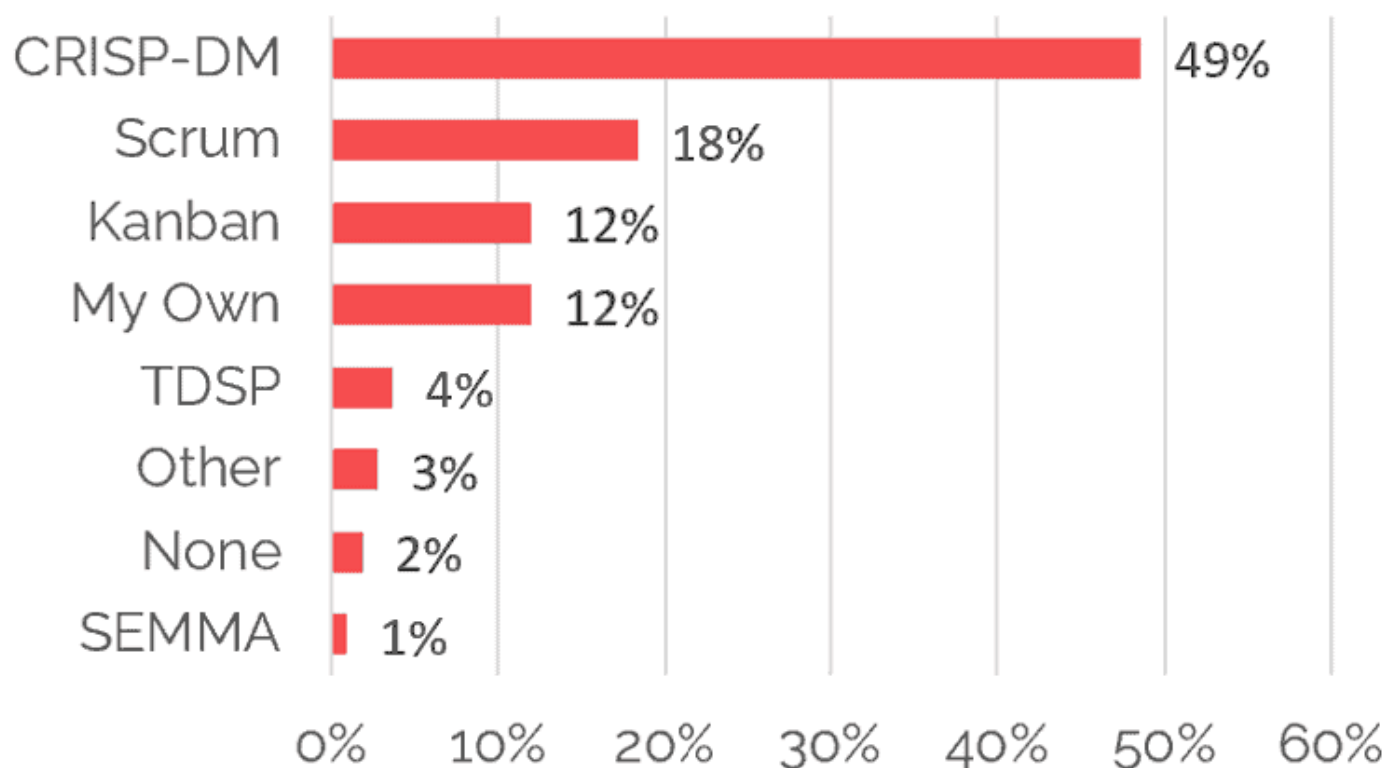
**Сергей Павлович Королёв**

Основоположник практической  
космонавтики

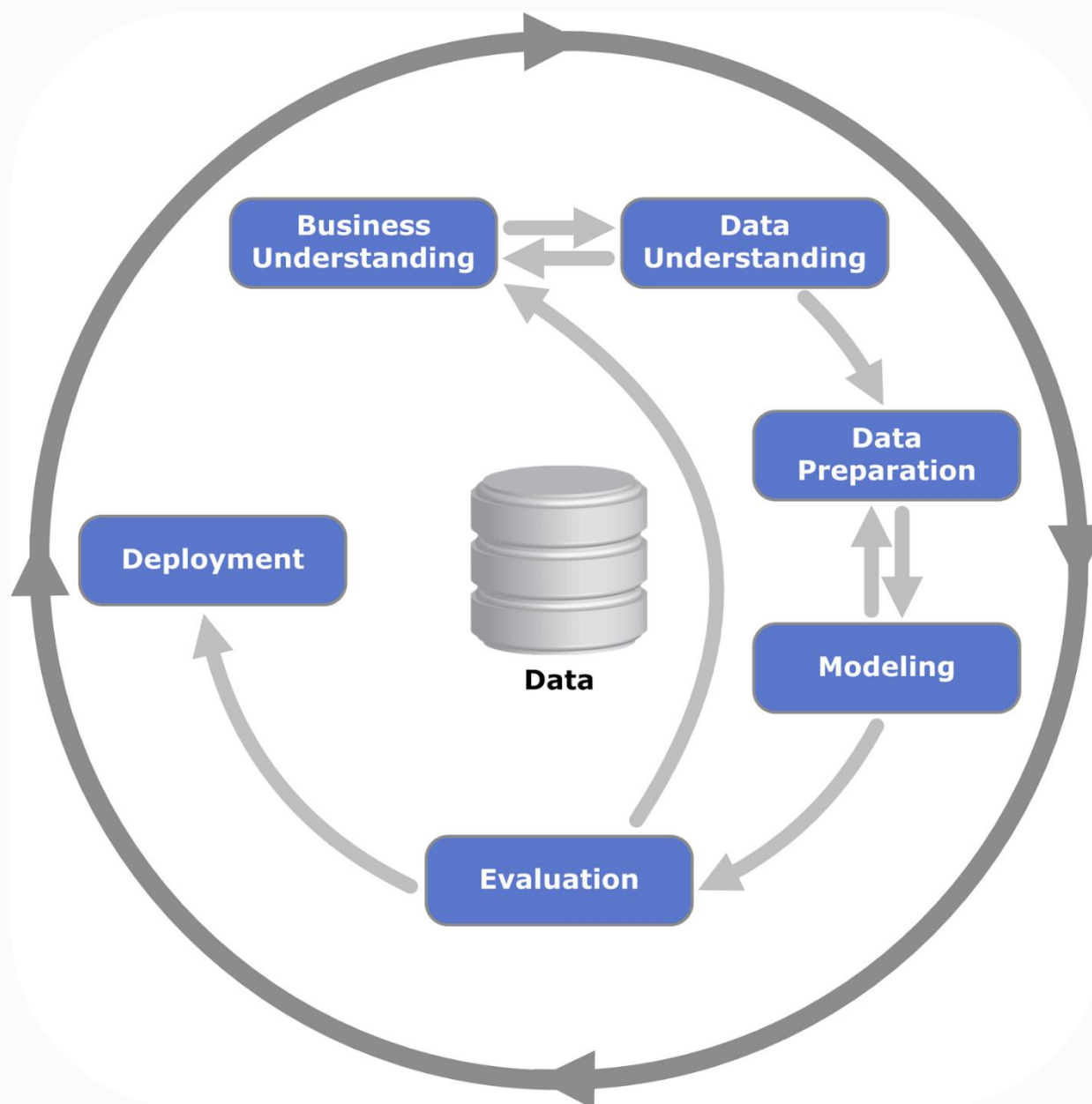
# Самая популярная методология

## datascience-pm.com Poll Results

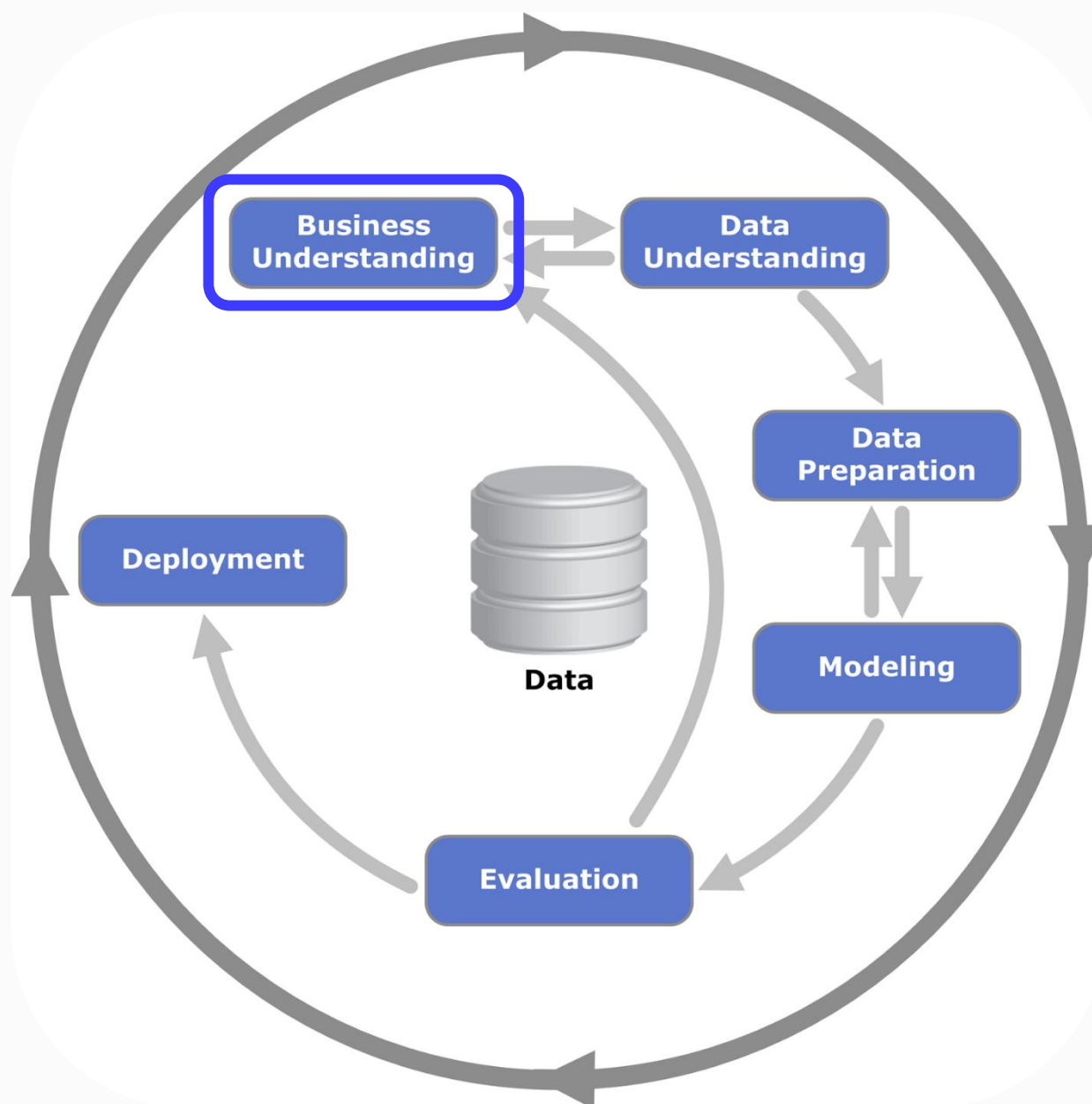
Which process do you most commonly use for data science projects?



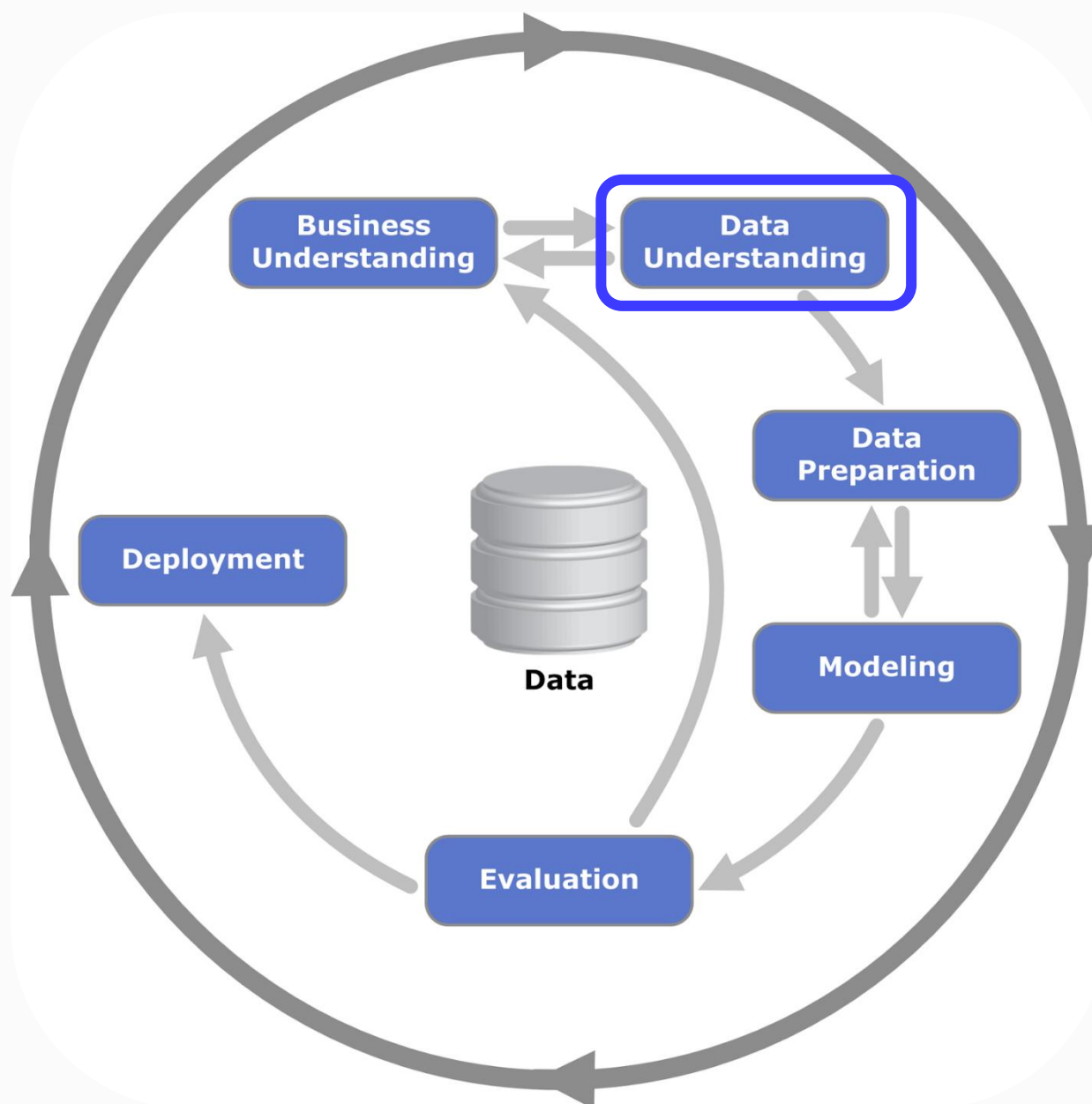
# Фазы CRISP-DM



# Фазы CRISP-DM



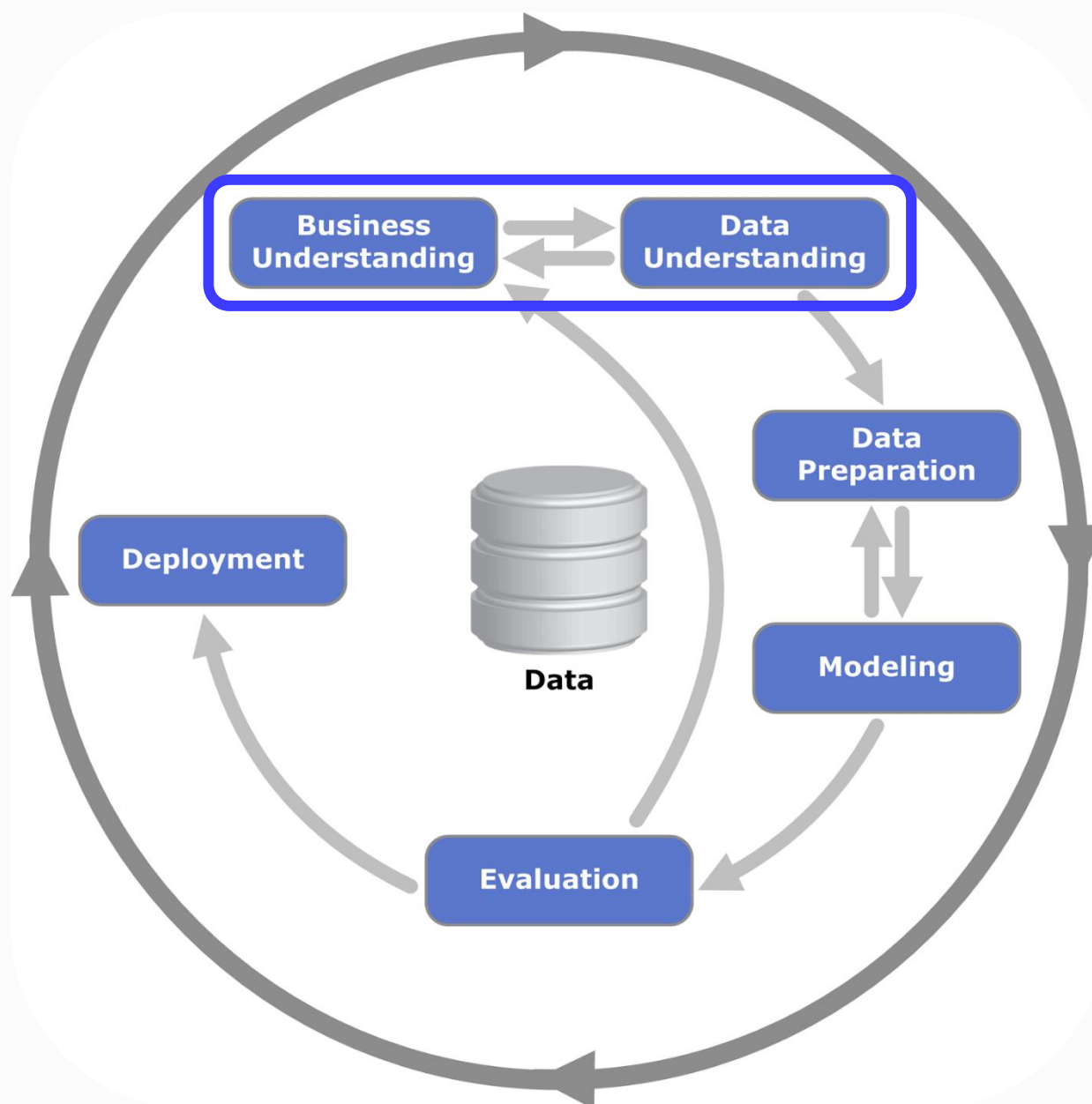
# Фазы CRISP-DM



# Data Understanding

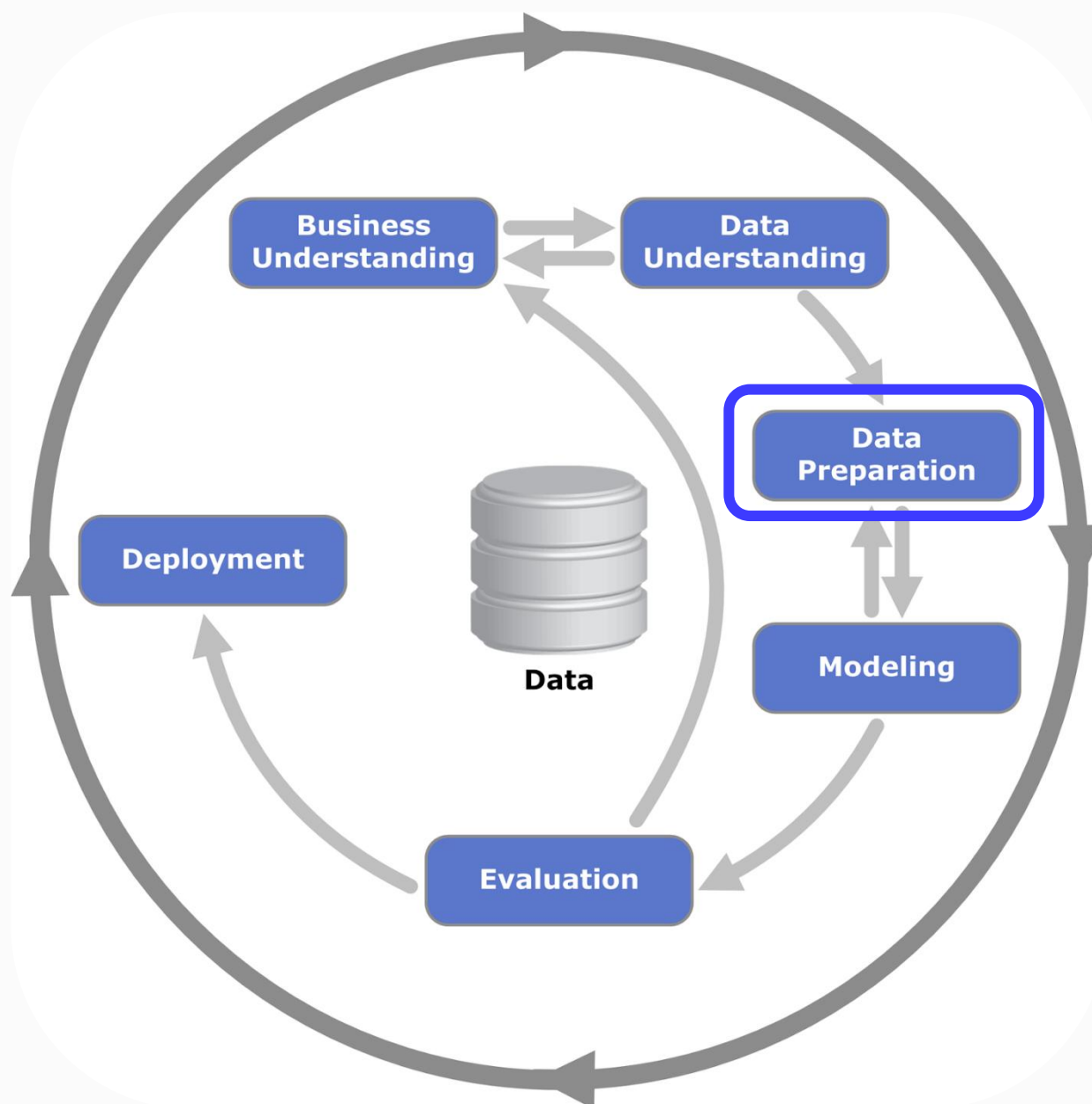
- 1 Собираем данные
- 2 Описываем данные
- 3 Исследуем данные
- 4 Проверяем качество данных

# Фазы CRISP-DM





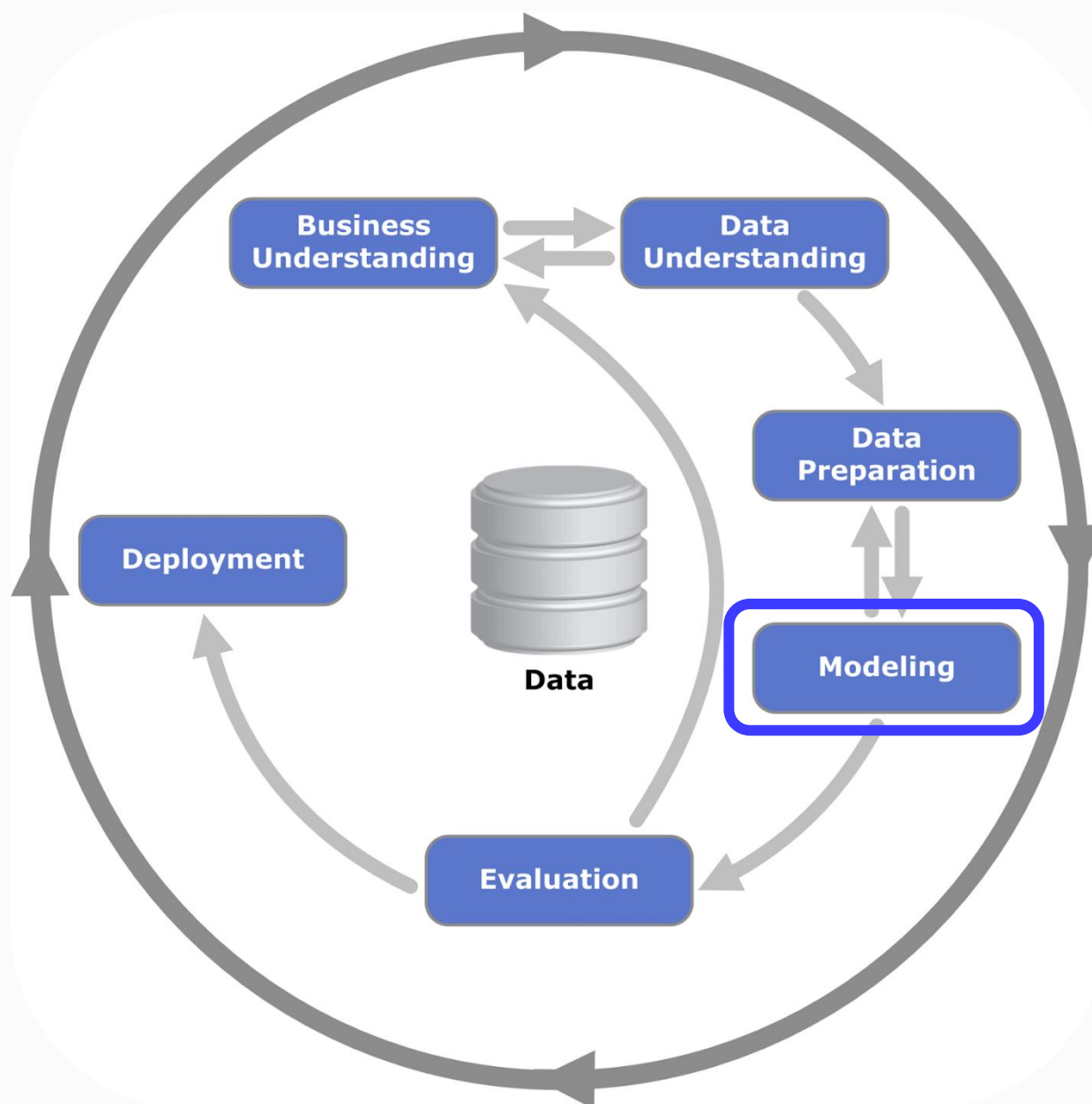
# Фазы CRISP-DM



# Data Preparation

- 1 Выбираем данные для анализа
- 2 Очищаем данные от мусора
- 3 Генерируем дополнительные данные
- 4 Объединяем данные в один датасет
- 5 Приводим данные к единому формату

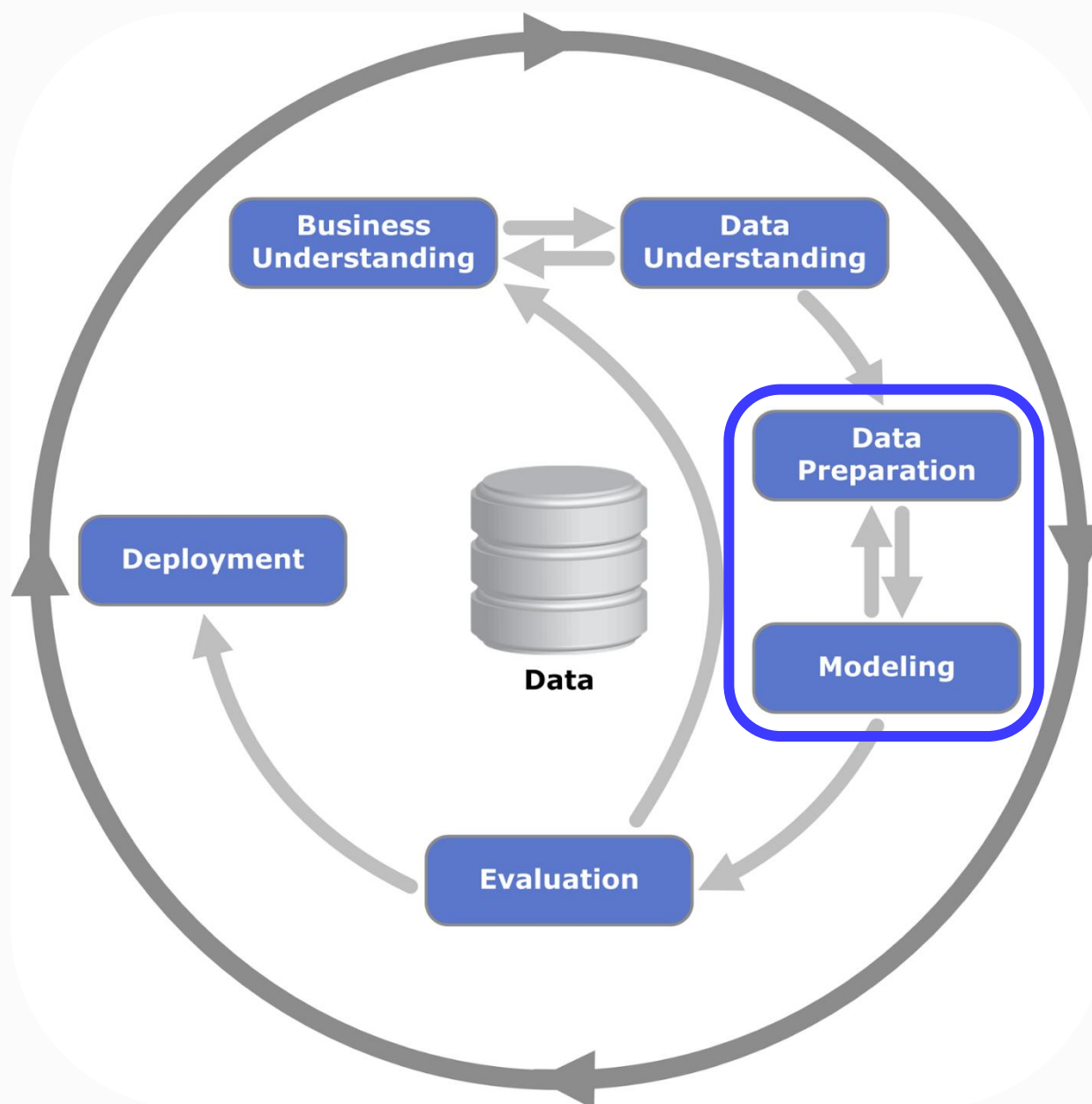
# Фазы CRISP-DM



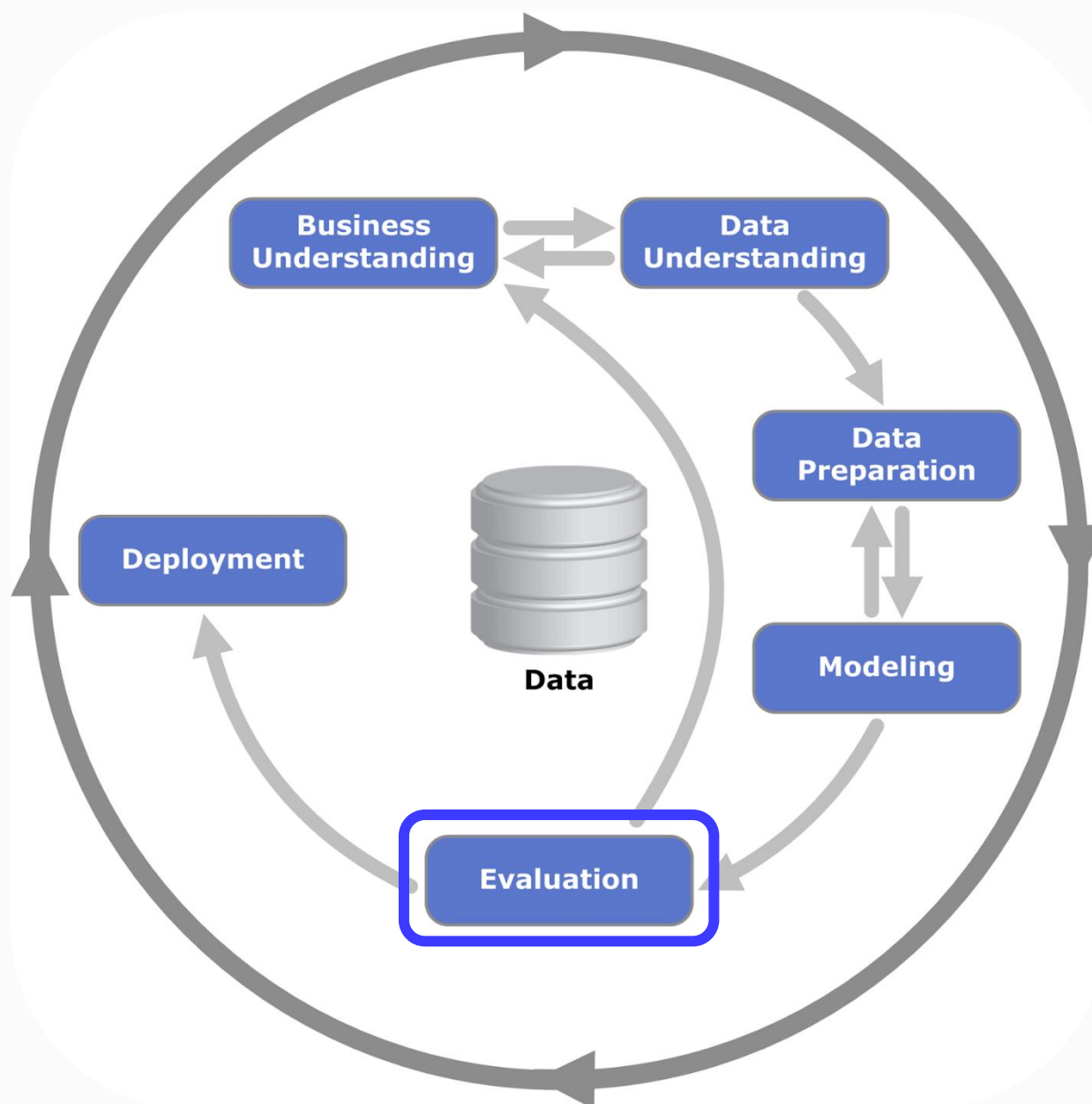
# Modeling

- 1 Выбираем наиболее подходящий алгоритм
- 2 Планируем эксперимент
- 3 Обучаем модель
- 4 Оцениваем результаты с технической точки зрения

# Фазы CRISP-DM



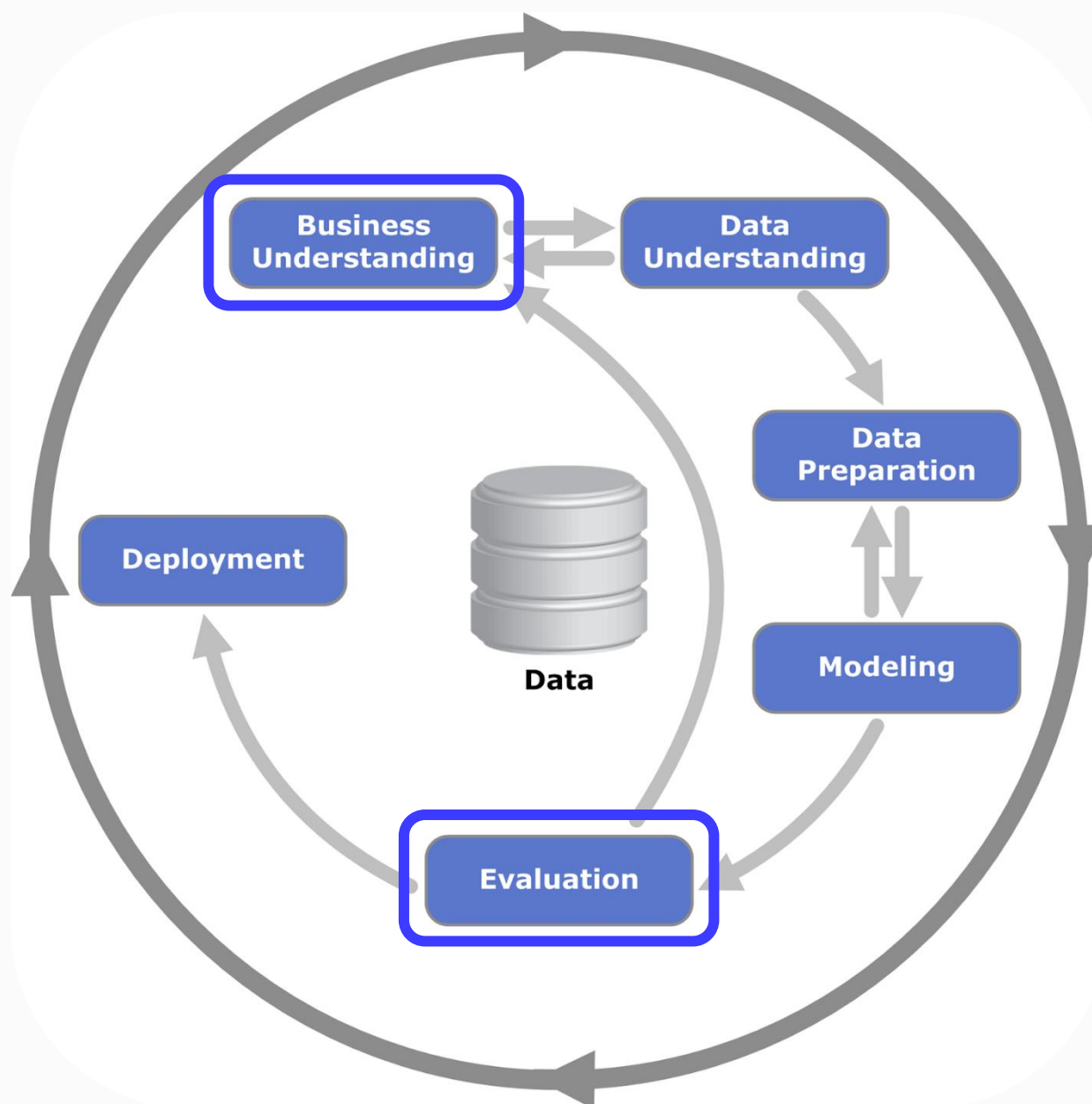
# Фазы CRISP-DM



# Evaluation

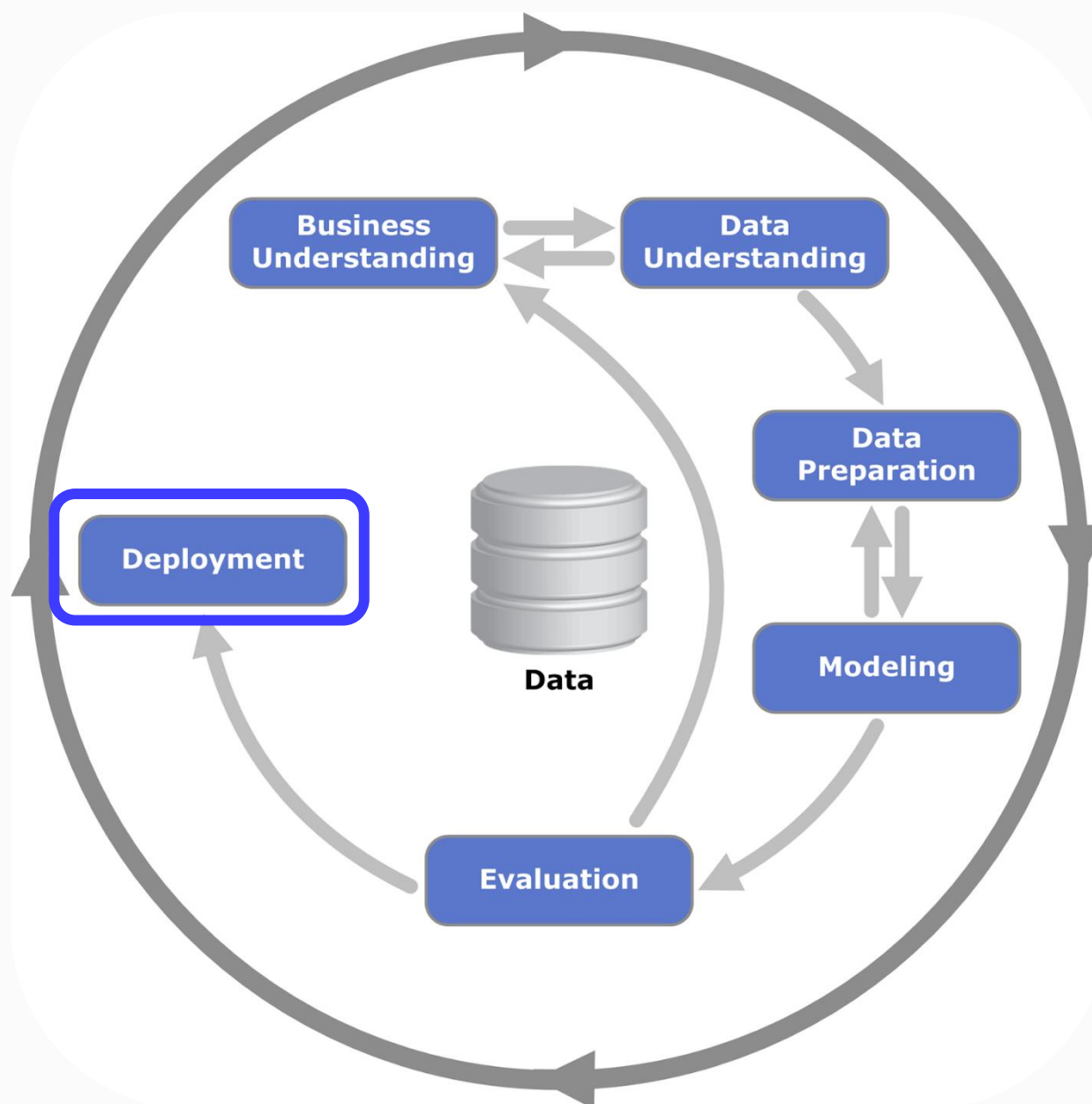
- 1 Оцениваем результаты с точки зрения бизнес-требований
- 2 Оцениваем процессы: что получилось хорошо, а что не очень
- 3 Определяем дальнейшее развитие проекта

# Фазы CRISP-DM





# Фазы CRISP-DM



# Deployment

- 1 Планируем процесс внедрения модели
- 2 Составляем план поддержки и мониторинга внедрения
- 3 Создаём подробный отчёт
- 4 Подводим итоги проекта