



WORKSHOP  
**Digital Data Analytics**  
Abordagens e Frameworks

Jorge da Costa Ferreira



WORKSHOP

# Digital Data Analytics

**QUE ABORDAGENS SE DEVE TER NUM PROJETO DE DATA ANALYTICS?**

Jorge da Costa Ferreira



## ABORDAGENS EM PROJETOS DE DATA MINING

O primeiro passo e toda a abordagem a realizar num projeto de Data Mining vai depender da metodologia aplicada. Para ajudar a normalizar os processos e fases de um estudo existem Frameworks estabelecidas.

### FRAMEWORKS:

**SEMMA** (desenvolvida pelo SAS institute)

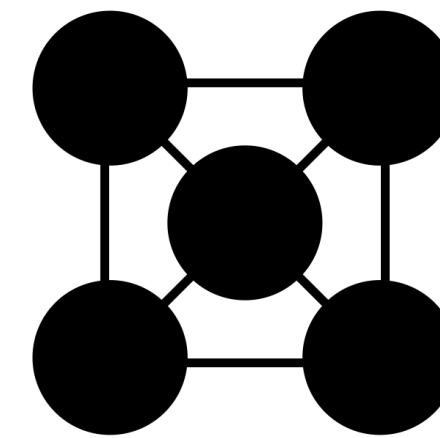


**CRISP-DM** (a minha favorita)

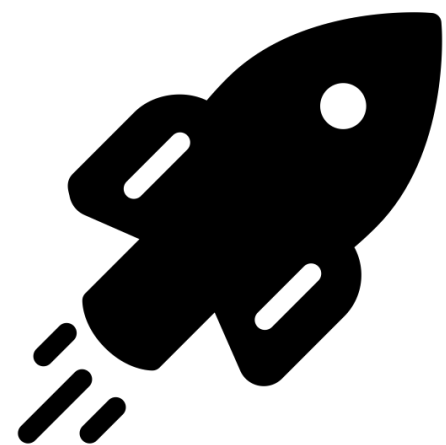
## SEMMA FRAMEWORK



**SAMPLE:** O processo inicia-se com amostragem de dados. O data set deve ser grande o suficiente para retornar informação relevante, mas pequeno para que seja eficiente



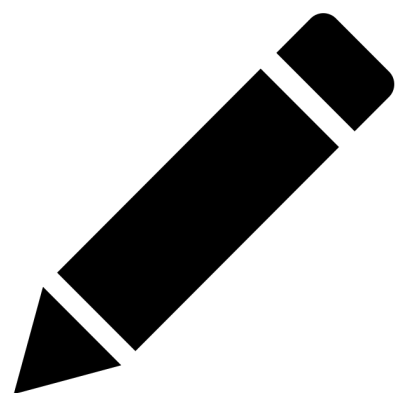
**MODEL:** Aplicação de modelos (técnicas de data mining) nas variáveis preparadas, com o objetivo de obter o conhecimento pretendido.



**EXPLORE:** Esta fase corresponde a compreender os dados através da identificação de relações esperadas ou não entre as variáveis; Este processo pode recorrer à mecanismos de visualização de dados



**ASSESS:** Avaliação da fiabilidade dos modelos criados



**MODIFY:** Criação e transformação de variáveis preparadas para a modelação

## SEMMA FRAMEWORK



○ SEMMA foca-se principalmente nas tarefas de modelação, **deixando de fora o know-how do negócio**

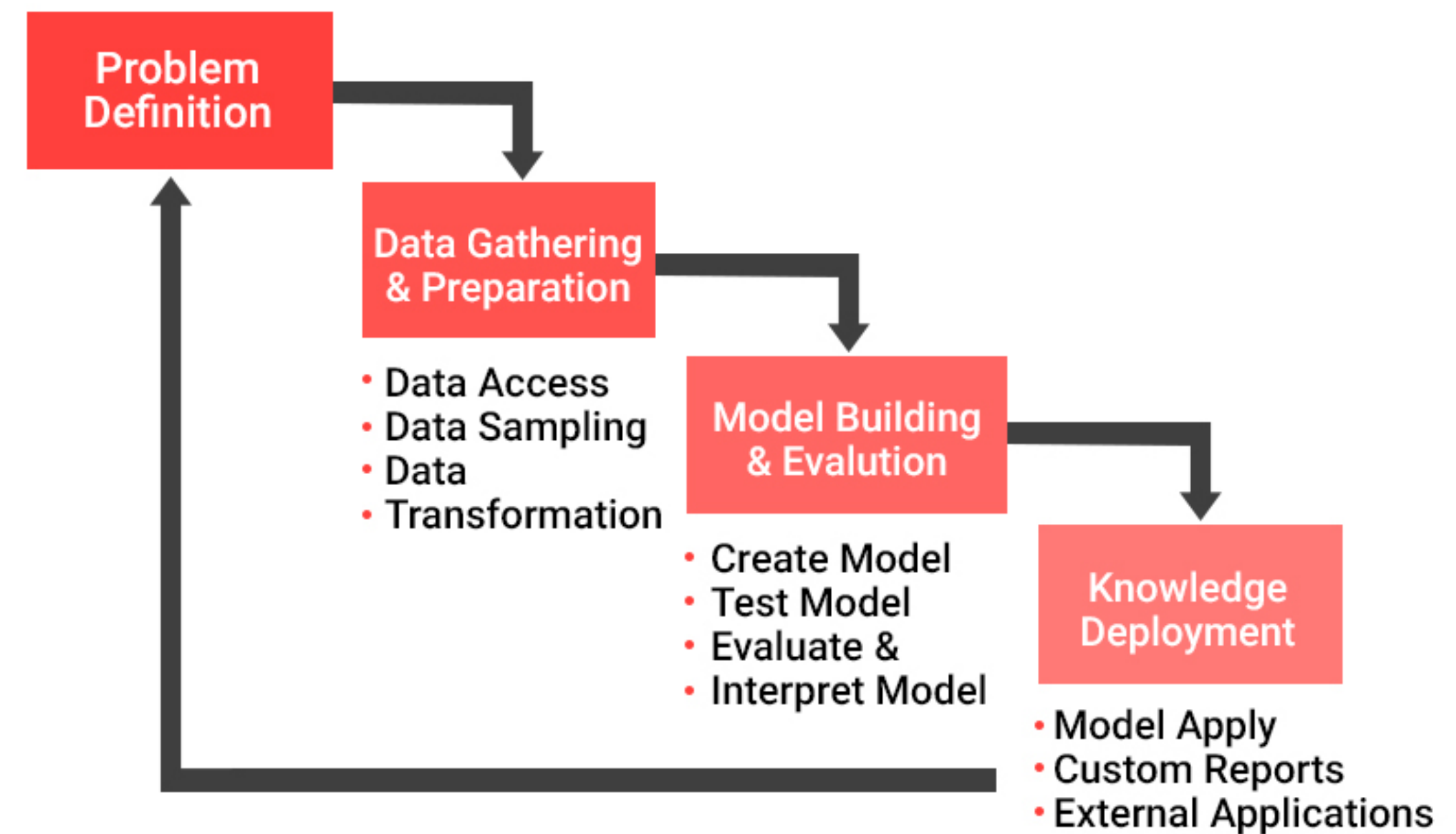
## CRISP-DM FRAMEWORK

## Cross Industry Standard Process for Data Mining

Uma visão mais ampla.

Pontos positivos:

- Standardização de processos
- Compreensão de negócio
- Acessibilidade de resultados



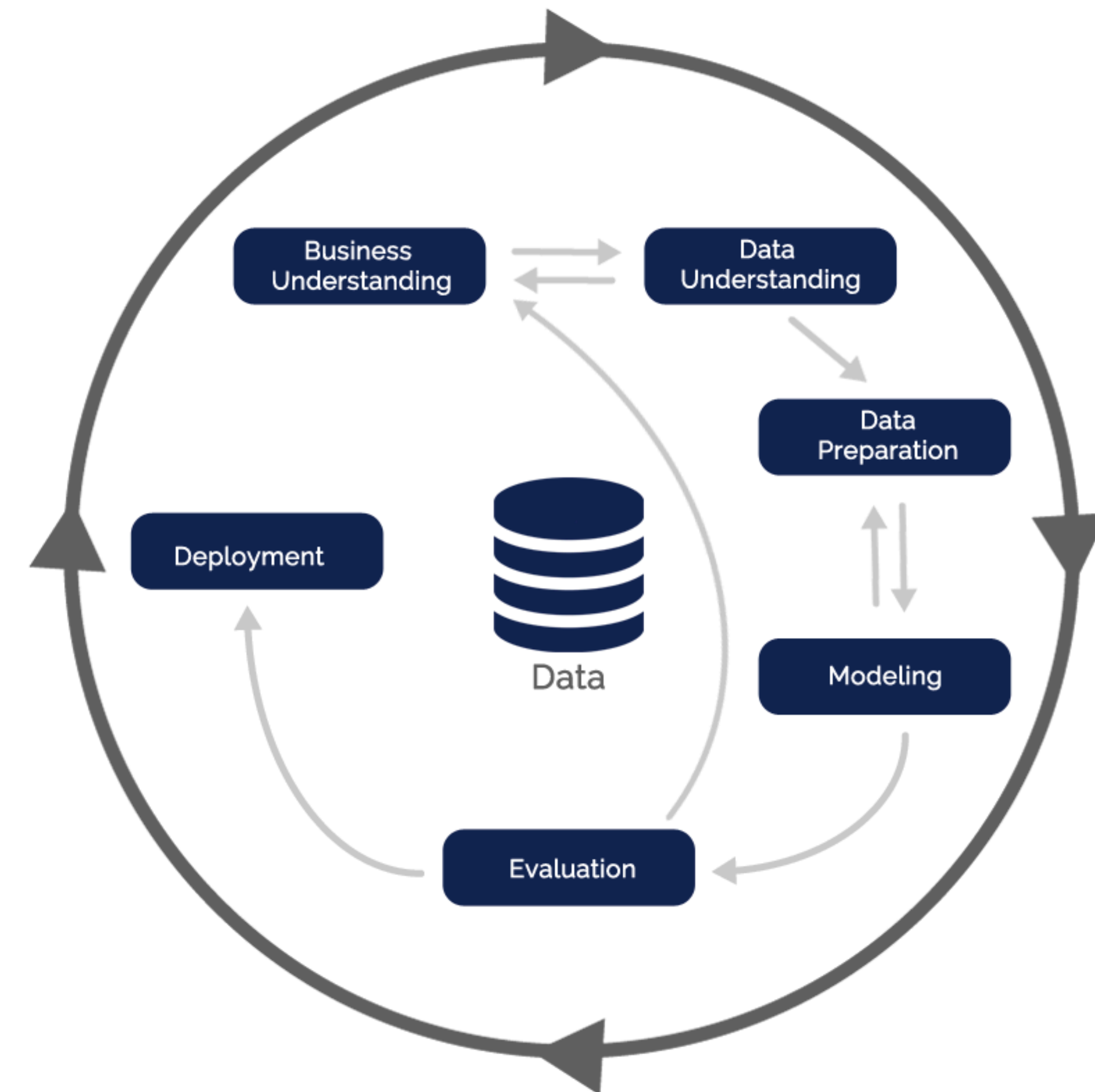
## CRISP-DM FRAMEWORK

**C**Ross Industry Standard Process for Data Mining

Uma visão mais ampla.

Pontos positivos:

- Standardização de processos
- Compreensão de negócio
- Acessibilidade de resultados



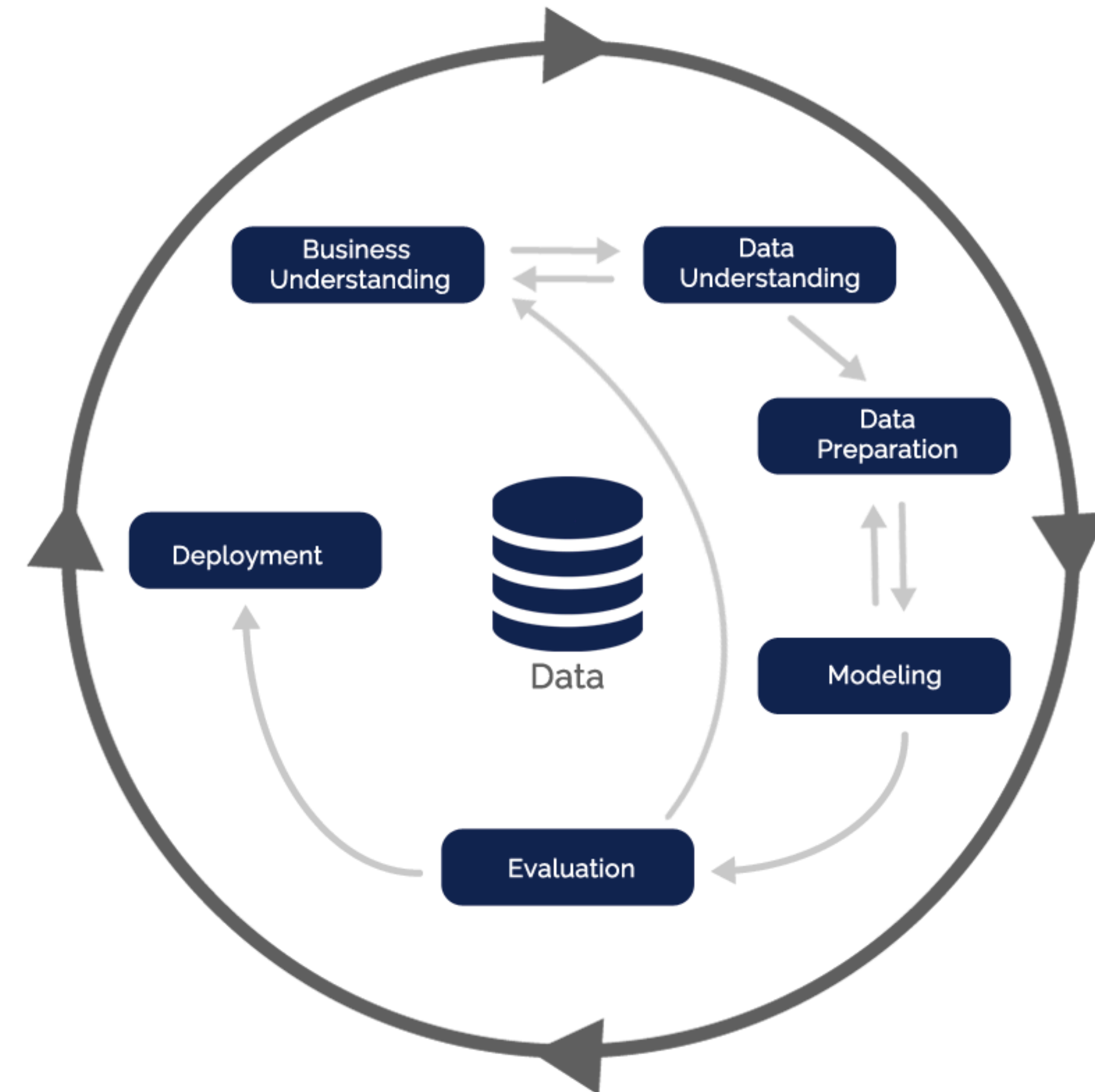
## CRISP-DM FRAMEWORK

**C**Ross Industry Standard Process for Data Mining

Uma visão mais ampla.

**Fase 1 - Business understanding:**

- Definição dos objetivos de negócio
- Definição dos objetivos do data mining
- Definição dos critérios de sucesso





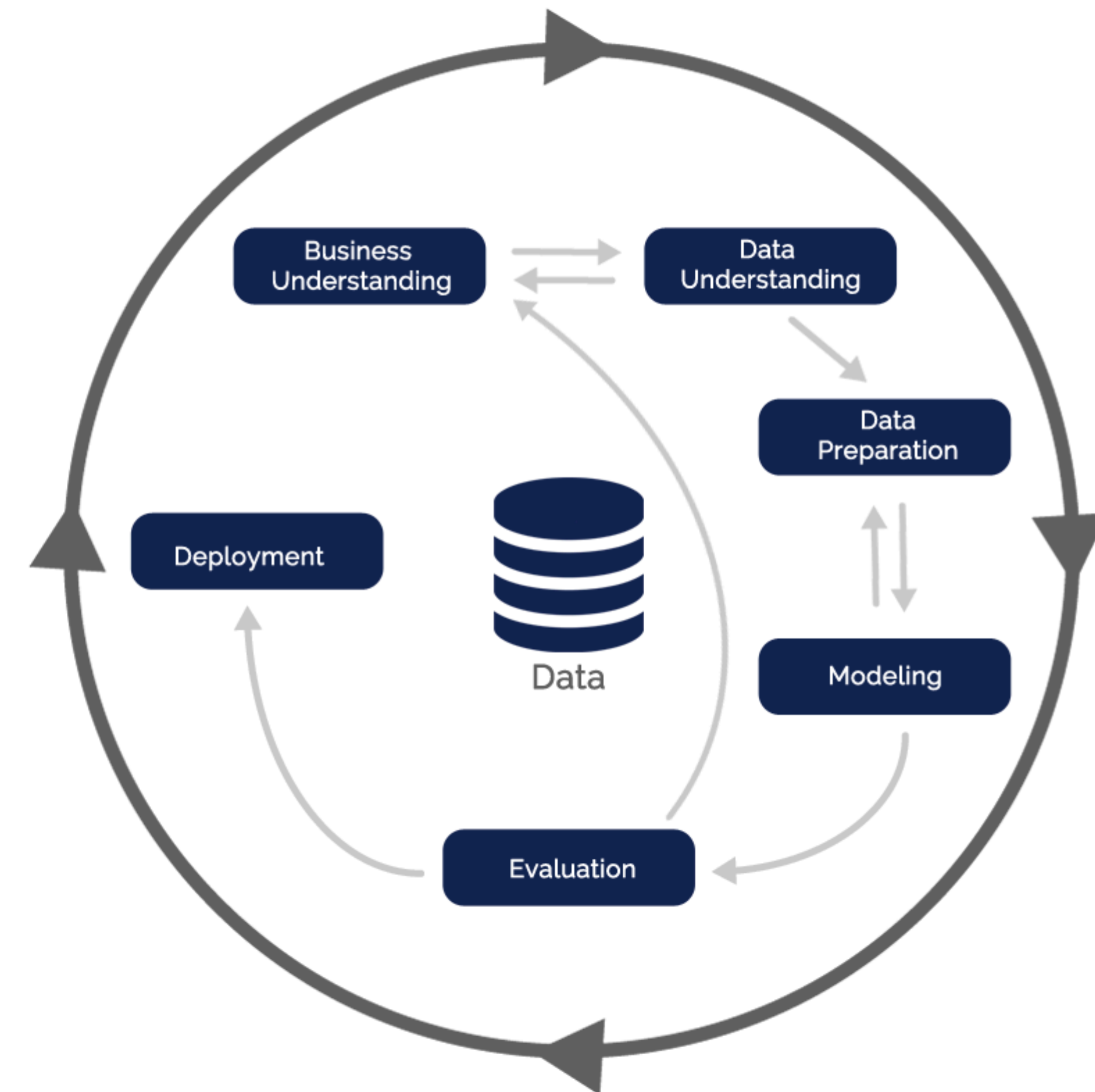
## CRISP-DM FRAMEWORK

**C**Ross Industry Standard Process for Data Mining

Uma visão mais ampla.

**Fase 2 - Data understanding:**

- Explorar os dados e validar qualidade
- Identificação de outliers



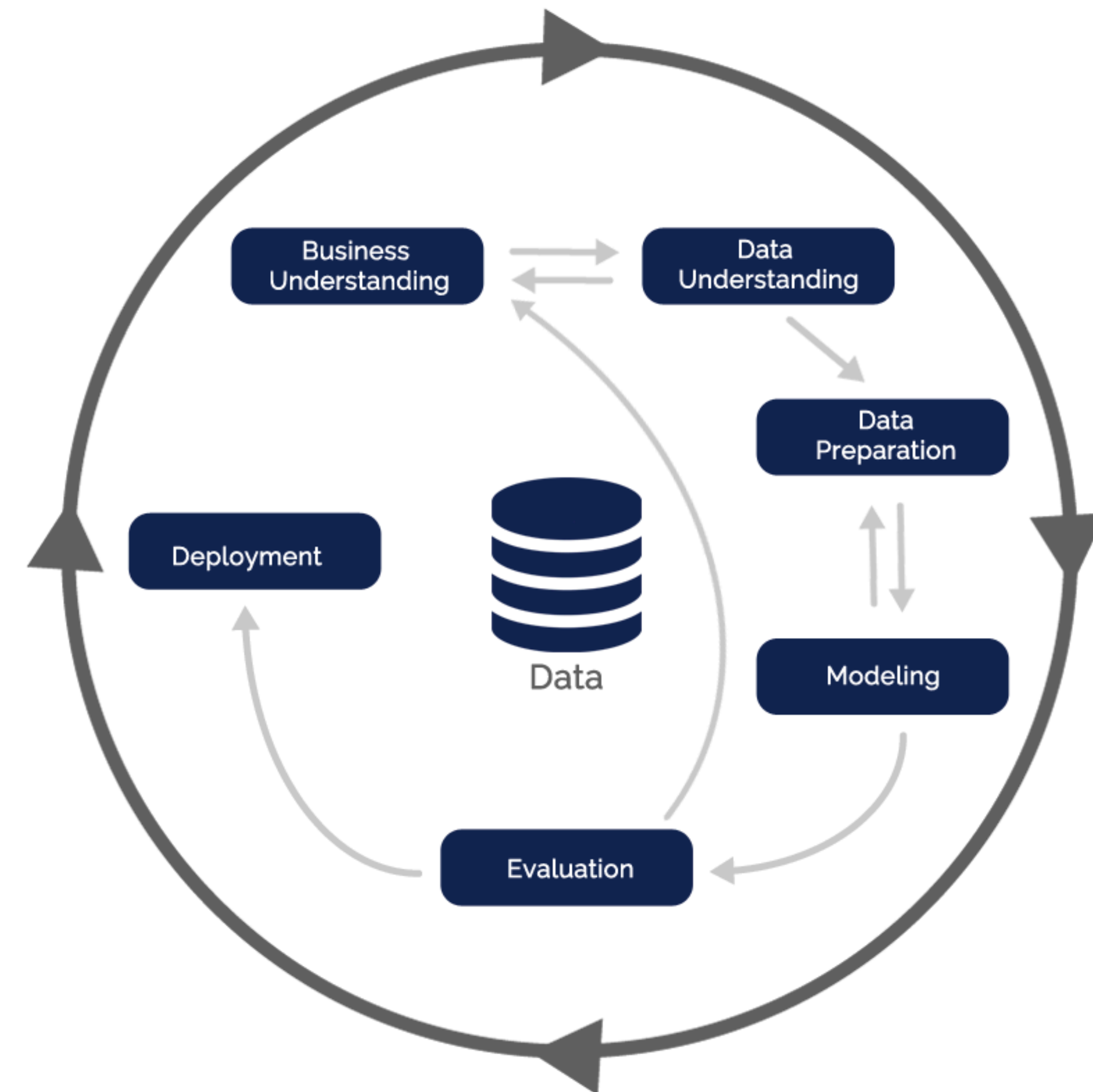
## CRISP-DM FRAMEWORK

**C**Ross Industry Standard Process for Data Mining

Uma visão mais ampla.

**Fase 3 - Data preparation:**

- Leva, normalmente, 90% do tempo do processo
- Coleção (múltiplas fontes)
- Avaliação Consolidação e limpeza
- Seleção de dados
- Transformação – data engineering



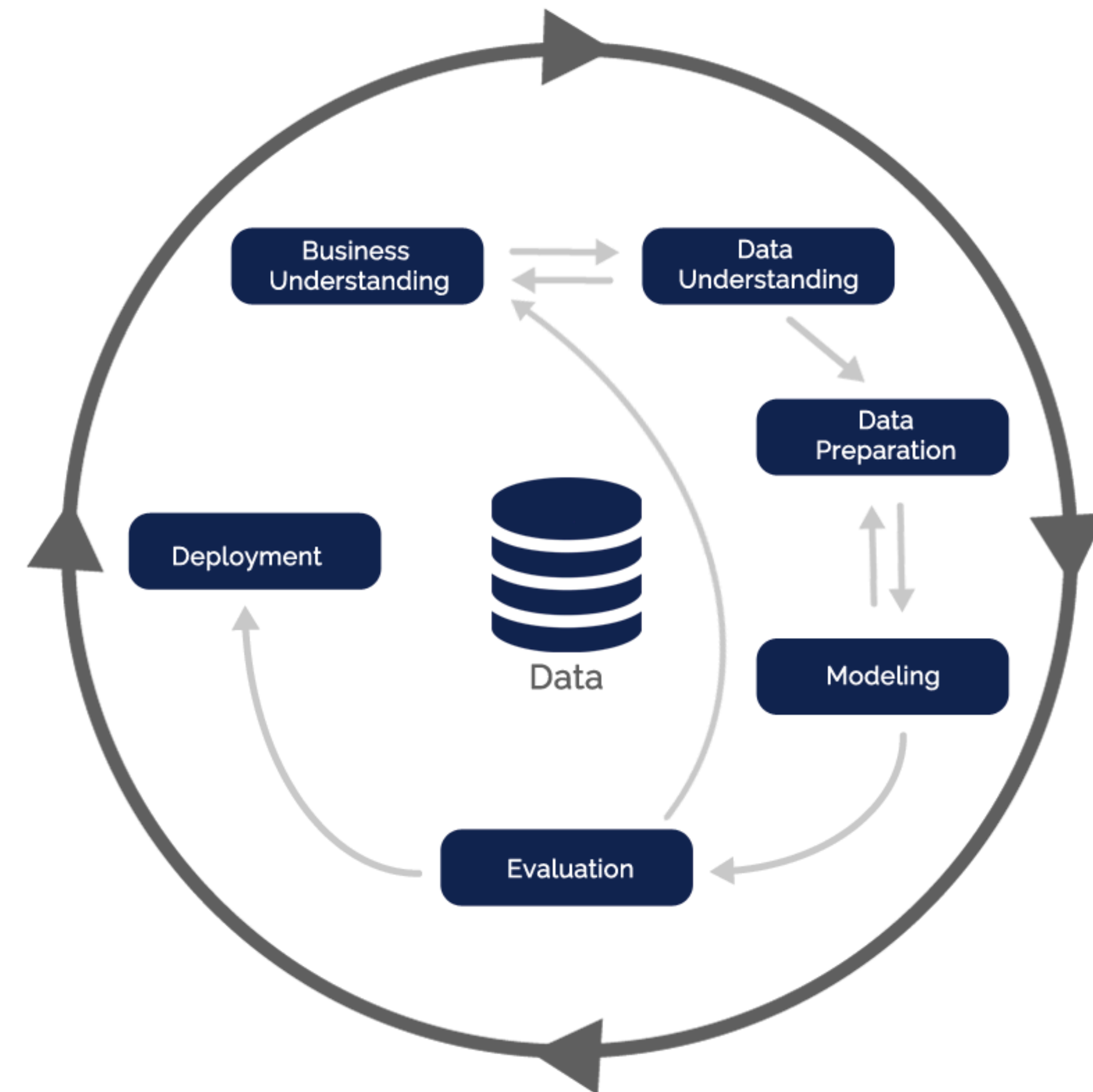
## CRISP-DM FRAMEWORK

Cross Industry Standard Process for Data Mining

Uma visão mais ampla.

**Fase 4 - Modeling:**

- Seleção do tipo modelo de acordo com o objetivo
- A modelação é um processo iterativo – diferente para problemas supervised e para unsupervised (vamos ver mais à frente)



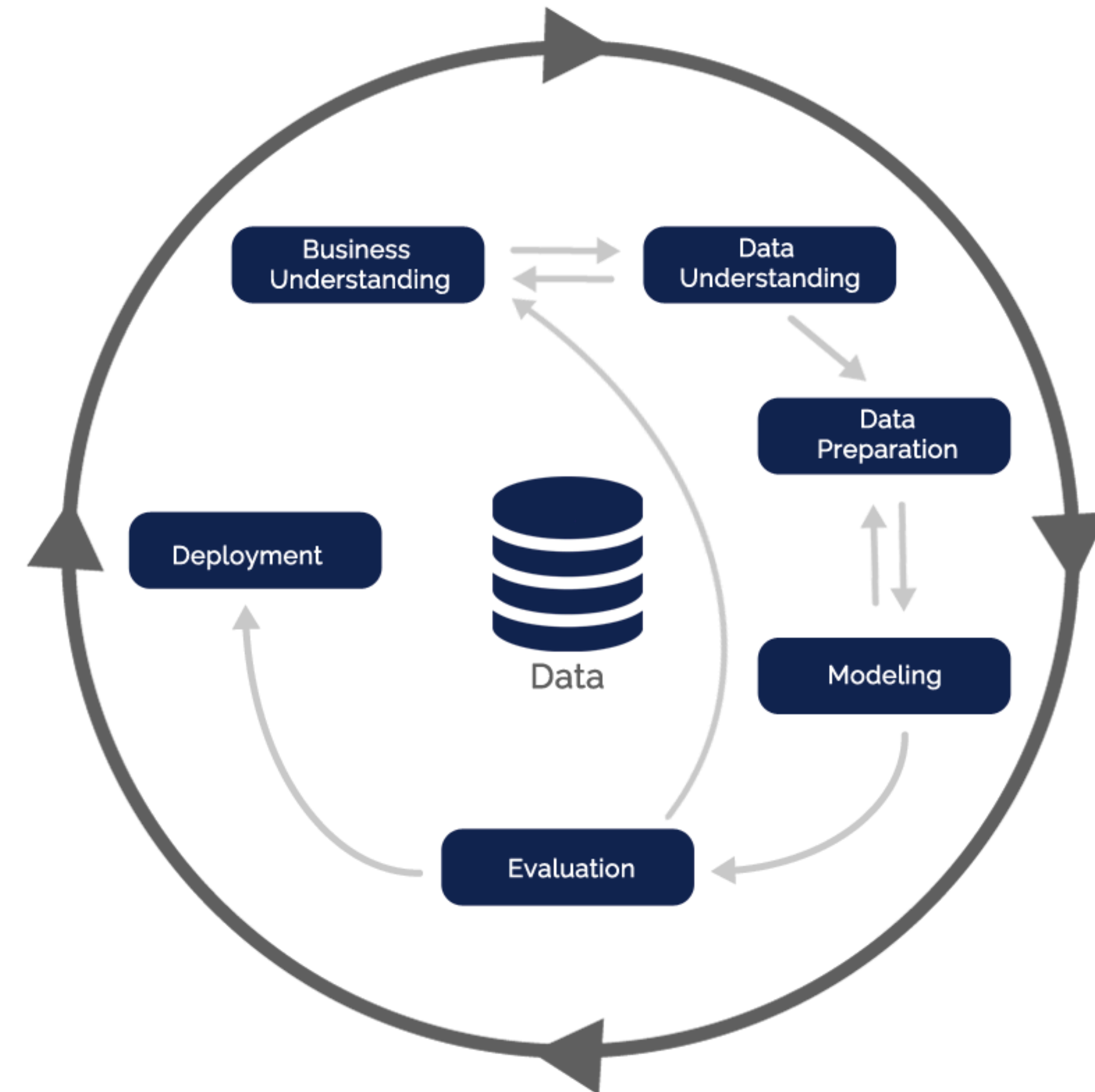
## CRISP-DM FRAMEWORK

Cross Industry Standard Process for Data Mining

Uma visão mais ampla.

**Fase 5 - Evaluation:**

- Testa o modelo em dados de teste para perceber quão bem representa a realidade
- Os métodos de avaliação do erro dependem do tipo de modelo (e.g. matriz de confusão em classificação e erro médio em regressões)





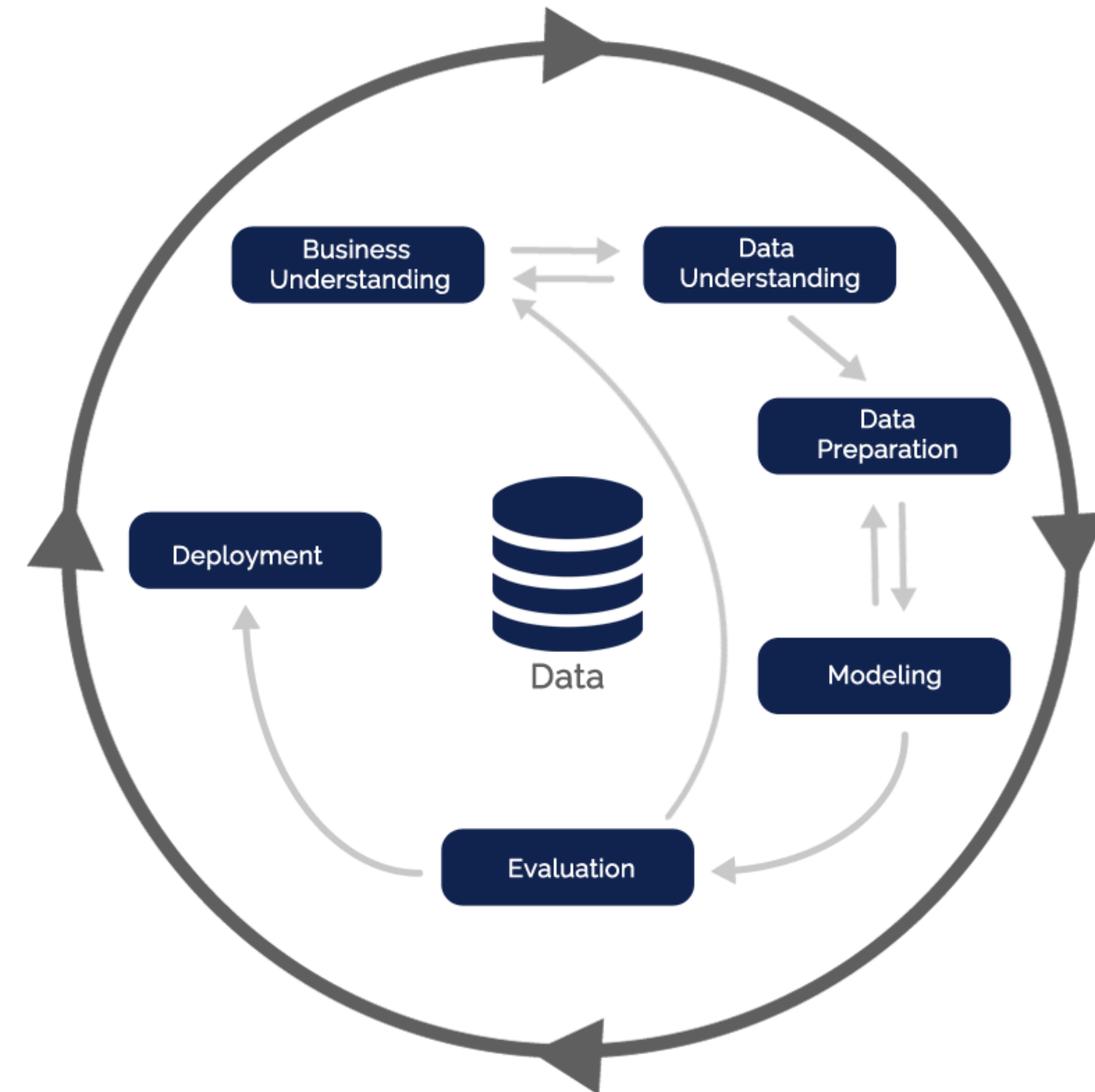
## CRISP-DM FRAMEWORK

Cross Industry Standard Process for Data Mining

Uma visão mais ampla.

**Fase 6 - Deployment:**

- Determinar como vão ser utilizados os resultados
- Quem precisa de os utilizar?
- Com que frequência?





WORKSHOP  
**Digital Data Analytics**  
Abordagens e Frameworks

Jorge da Costa Ferreira