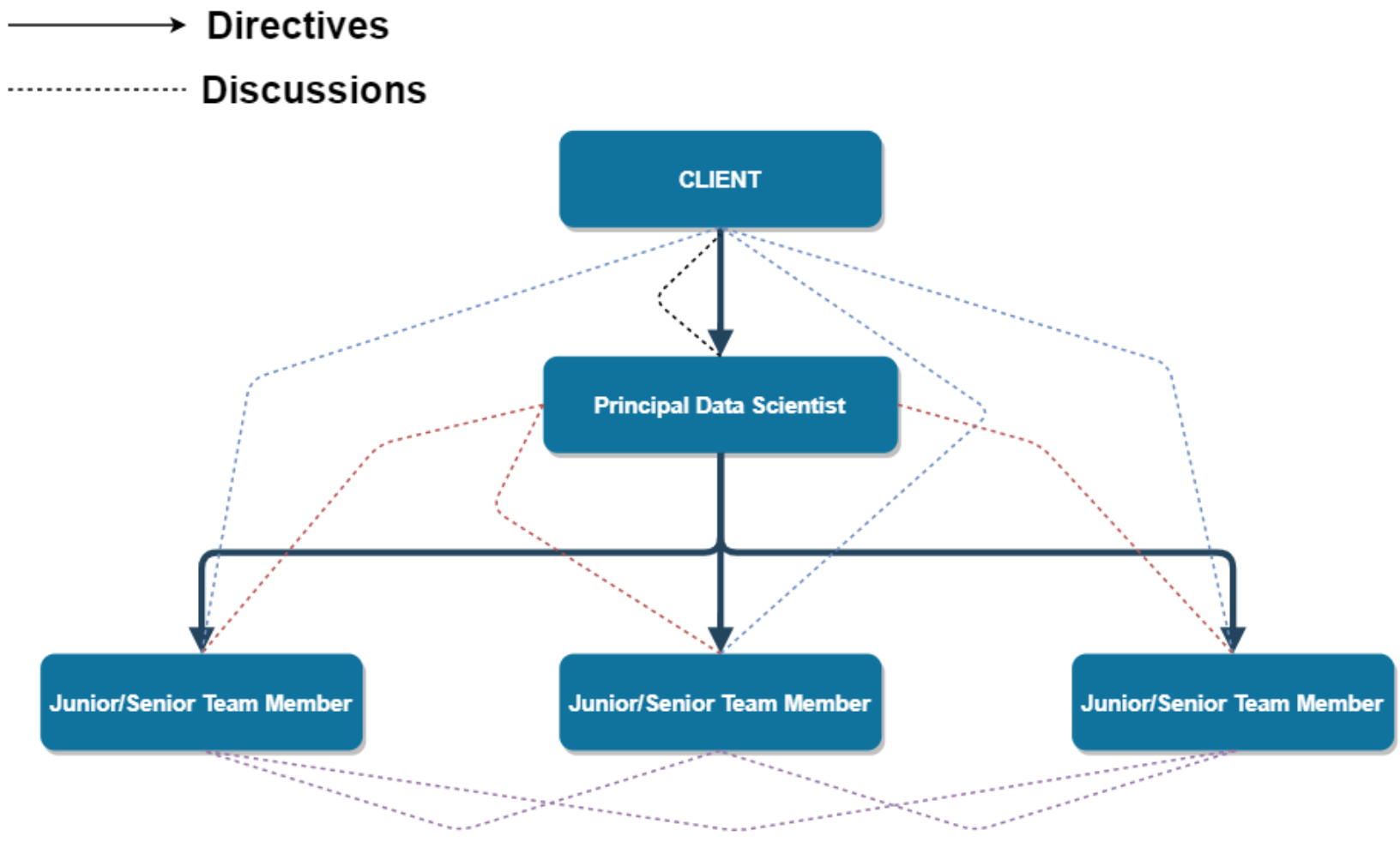
Project Management and Data Science

Ryan Miller

Organizational Structure

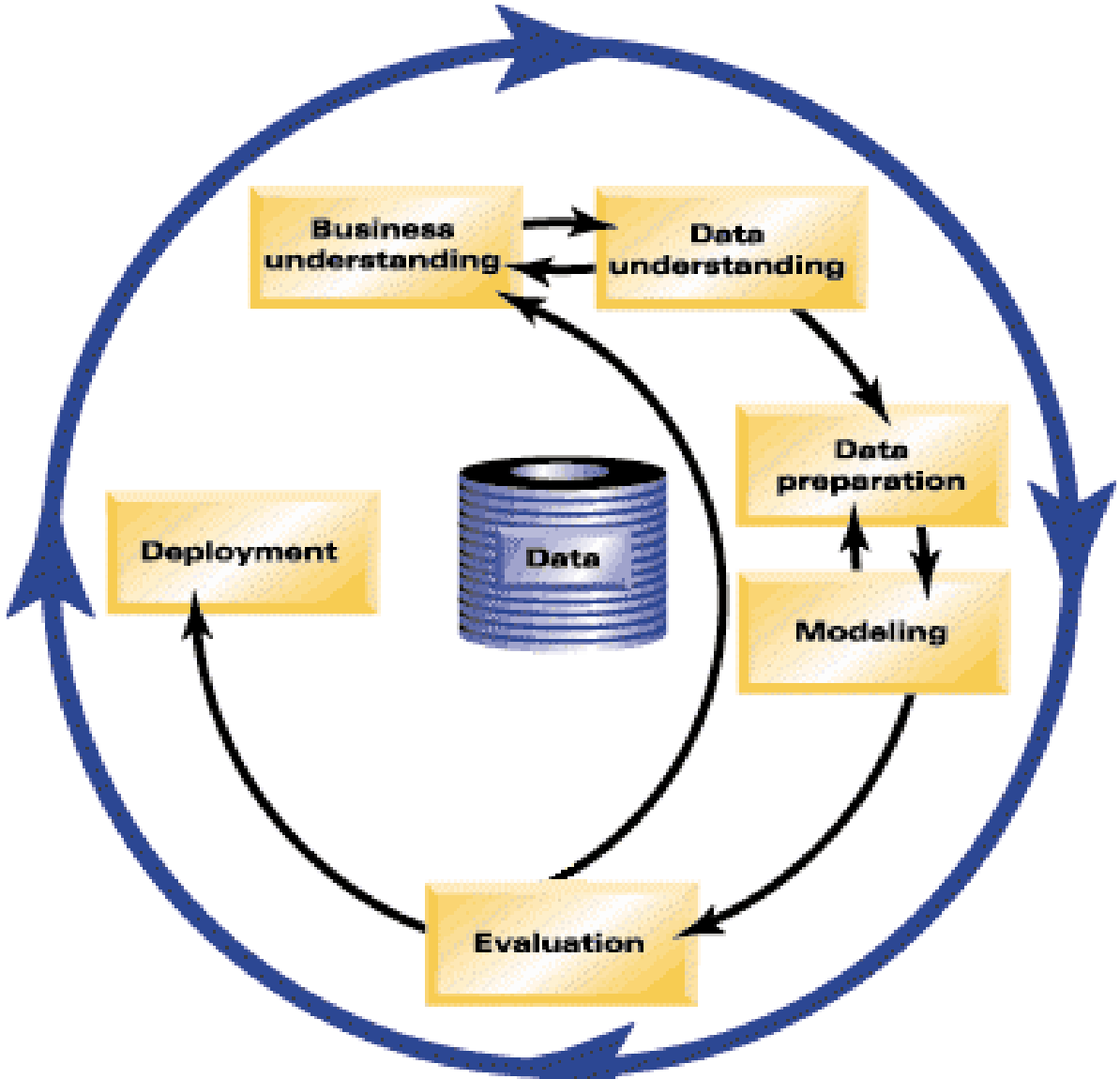


Data Science “pipelines” and “life-cycles”

- The data science work-flow takes place within the organizational framework
- While “Data science” is oft used, its definitions vary
  - Click here to see some examples of this

CRISP-DM

We'll adopt the Cross-Industry Standard Process for Data Mining (CRISP-DM) model:



Phase 1 - Business Understanding

Tasks:

1. Gather necessary background information
2. Document specific specific objectives
3. Determine success criteria for the project

Each of these tasks should be undertaken in coordination with the client

Phase 1 - Objective vs. Subjective Success Criteria

- You should have a mixture of objective and subjective success criteria
  - Objective = “Increase the time visiters spend on the destinations page by 10%”
  - Subjective = “Identify customer clusters”

Phase 2 - Data Understanding

Tasks:

1. Describing the data
2. Exploring the data
3. Verifying data quality

These tasks should be carried out at the team level (and cross-referenced with the principal and client if necessary)

Phase 3 - Data Preparation

Tasks:

- Merging/joining (ie: `left_join`)
- Selecting relevant subsets (ie: `filter`)
- Aggregating records (ie: `group_by` and `summarize`)
- Deriving new attributes (ie: `mutate`)
- Handling missing data (ie: `complete.cases` or `knnImput`/`rfImpute`)

These tasks should be carried out at the team level and cross-referenced by the principal (they are seldom relevant to the client at this point)

Phase 4 - Modeling

Tasks:

- Selecting a model
- Evaluating the "goodness" of a model
- Building the model
- Note: you may replace "model" with "product" in some applications

This phase is highly non-linear, it should be carried out at the team level and cross-referenced by the principal (and possibly with the client depending on their level of technical proficiency)

Phase 5 - Evaluation

Tasks:

1. Consider your model/product in regards to the business success criteria you came up with in Phase 1
2. Formalize your findings

These tasks should be undertaken in coordination with the client and principal

Phase 6 - Deployment

Tasks:

1. Deliver your model/product to the client
2. Complete wrap-up tasks (ie: technical report, etc.)

Practice