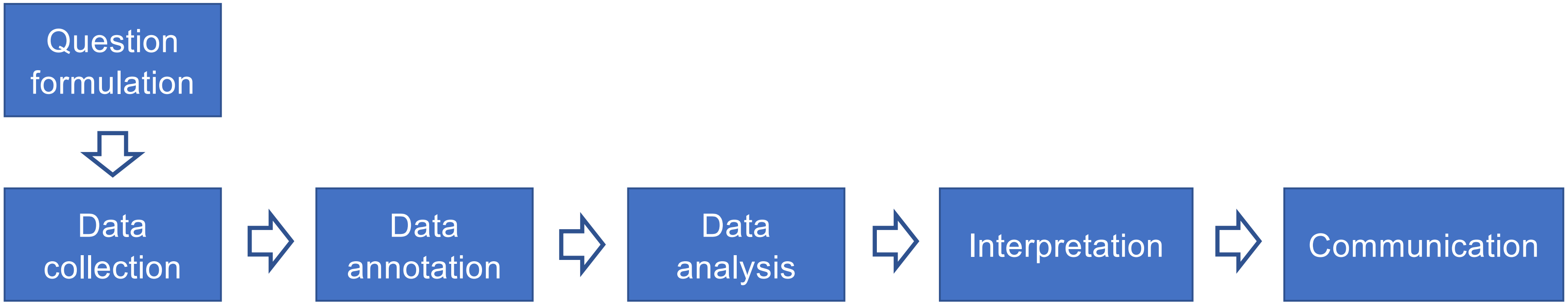
Unit 4: Question Formulation

Why does a good question look like?

Lesson 35

Derek Ruths

Data science project phases



Overview of unit

Objectives:

- Understand why question formulation is a non-trivial… often the most important part of the data science project
- Know techniques for approaching question formulation

1. Why question formulation?
2. What does a good question look like?
3. Case study
4. Question formulation techniques

Lesson overview

Objectives

- Understand what makes for a good data science question

Outline

- The stakeholder, the use case
- Adding the details

person who is not you and is interested in the question

The stakeholder, the use case

Someone is asking this question for a reason…

- chances are it’s not you (or just you)
- chances are you’re not going to be the (only) one to use the information

Examples from my own experience:

- Political online violence in developing countries (NDI)
- Measuring severe acute malnutrition rates in WFP clinics (WFP Nutrition)
- Detecting hate speech in Reddit forums

Adding details

Details of the question must be as precise as possible

- "Political online violence in developing countries" ← original question
- "Twitter-based violence against politically-active people in developing countries" ← get more specific
- "Directed abusive tweets against politically-active people in developing countries"
- "Directed abusive tweets against non-politician, politically-active accounts in developing countries"
- "Directed abusive tweets against non-politician, politically-active accounts that self-identify as citizens in Indonesia, Colombia, and Kenya"

conversation based changes

What question gives data?

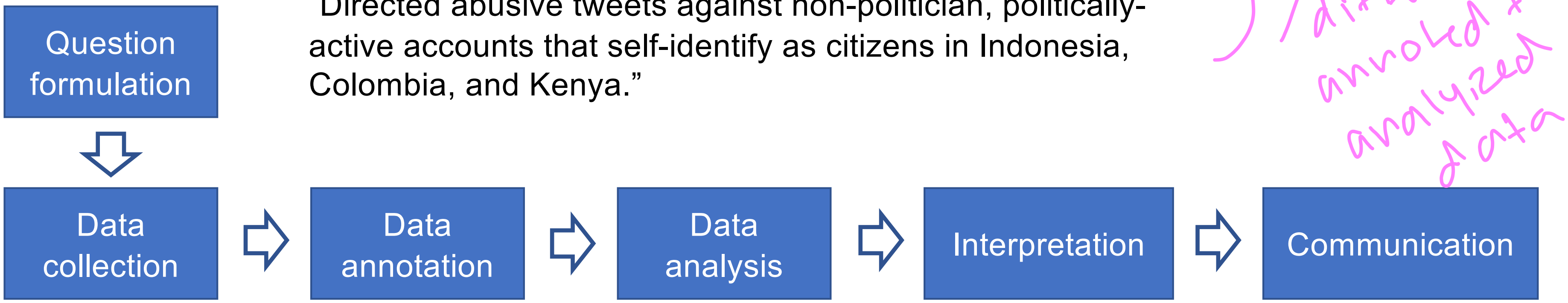
Why does getting the question right matter?

"Online violence against in politics"

"Abusive retweets to politician account messages in Indonesia, Colombia, and Kenya"

"Directed abusive tweets against non-politician, politically-active accounts that self-identify as citizens in Indonesia, Colombia, and Kenya."

all of these involve different data / differently annoted + analyzed data



Components of a good question

Detailed use case:

- “We’re interested in raising the visibility of online violence that keeps people from participating in political movements. We’ll use these results to make a quantitative case that online violence is a substantial part of people’s lived experience of social media and informs the way they engage in politics.”

Detailed question:

- “Directed abusive tweets against non-politician, politically-active accounts that self-identify as citizens in Indonesia, Colombia, and Kenya”

Lesson wrap-up

Takeaways

- A good question is grounded in clear, well-understood use cases
- A good question provides detailed characterizations of each part of the question.

Up next

- Techniques for getting to a good question formalization