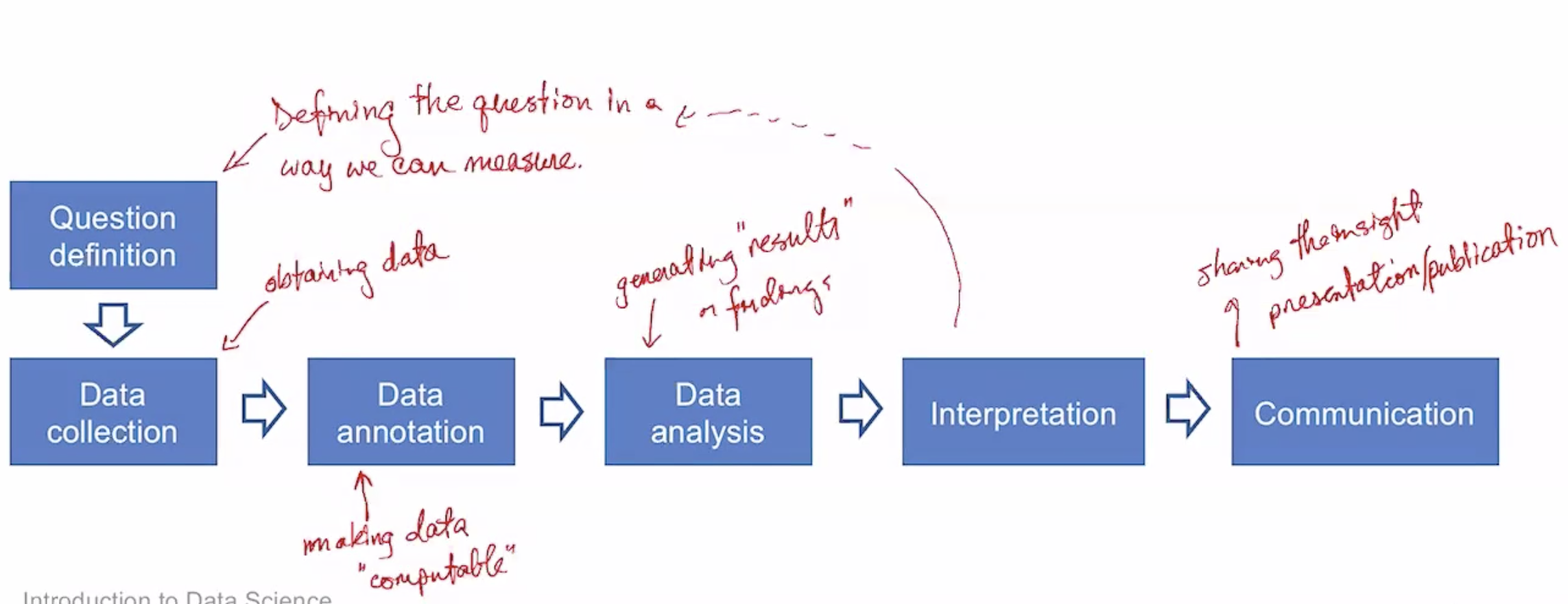
[Definitions of Data Science]

* **Data:** A collection of recorded observations.
* **Quantitative data:** Data can be represented as numbers.
* **Feature:** A (possibly inferred) attribute of data; usually quantitative in DS
* **Proxy:** A feature that approximates a particular thing we want to quantify

[Data Science Project Phase]



1. Data Collection:

* Decide what to collect
* Obtaining the raw data
* Identifying anomalies (by plots)
* Correcting issues in the data
* Standardizing the structure of the data e.g. database, csv

2. Data Annotation: Applying/inferring features that will be used for analysis

* Human annotation (Expert annotation / Crowd sourced annotation)
* Machine learning

3. Data Analysis: The creation of statistical models and visualizations using existing methods and existing, (relatively) clean quantitative data