

Data Science Essentials

Principles of Data Science

Data science is about using data to make decisions that drive actions. The goal of data science is to use data analytic thinking to:

- Replace intuition with data driven analytical decisions
- Transform raw data to valuable asset
- Increase pace of action

Data science involves:

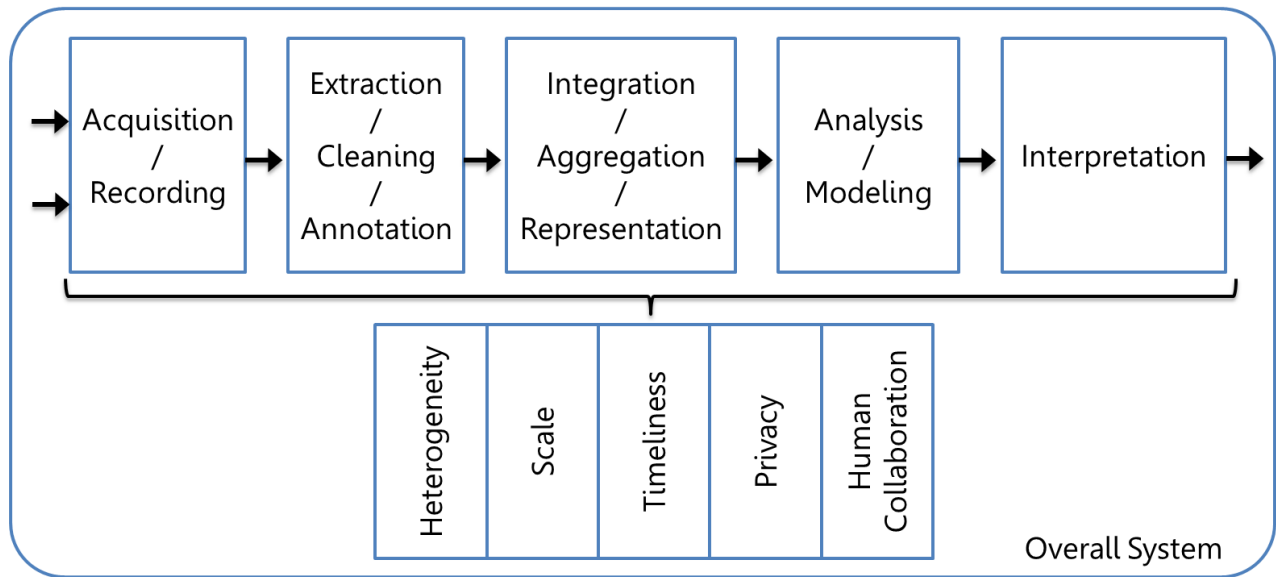
- Finding data
- Acquiring data
- Cleaning and transforming data
- Understanding relationships in data
- Delivering value from data

The data science process includes the following activities:

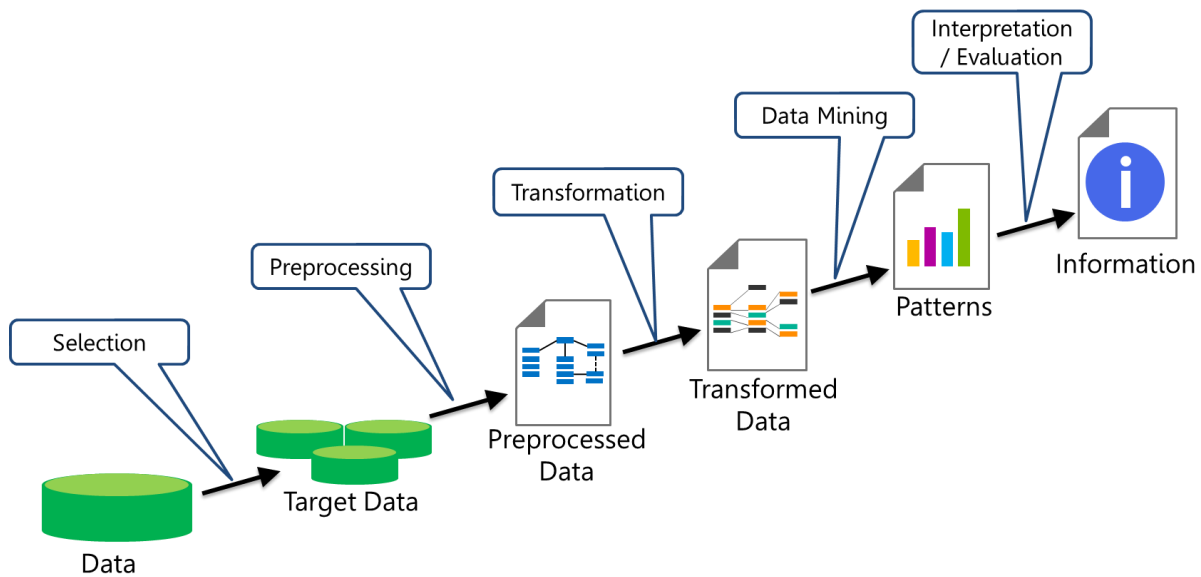
1. Data selection.
2. Preprocessing.
3. Transformation.
4. Data Mining.
5. Interpretation and evaluation.

It is an iterative process in which some, or all, steps may be repeated.

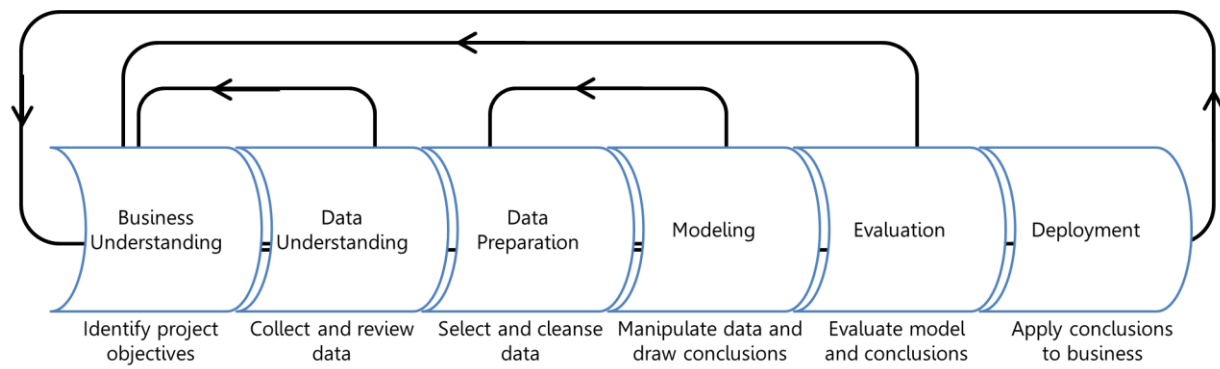
Some attempts to formalize the process include:



The Computing Community Consortium Big Data Whitepaper (2012).
<http://www.cra.org/ccc/files/docs/init/bigdatawhitepaper.pdf>



The Knowledge Discovery in Databases (KDD) process (1997)
<http://www.aaai.org/ojs/index.php/aimagazine/article/view/1230>



The Cross Industry Standard Process for Data Mining (CRISP-DM) (2000)

(http://spss.ch/upload/1107356429_CrispDM1.0.pdf)