A Network Tour of

Data Science

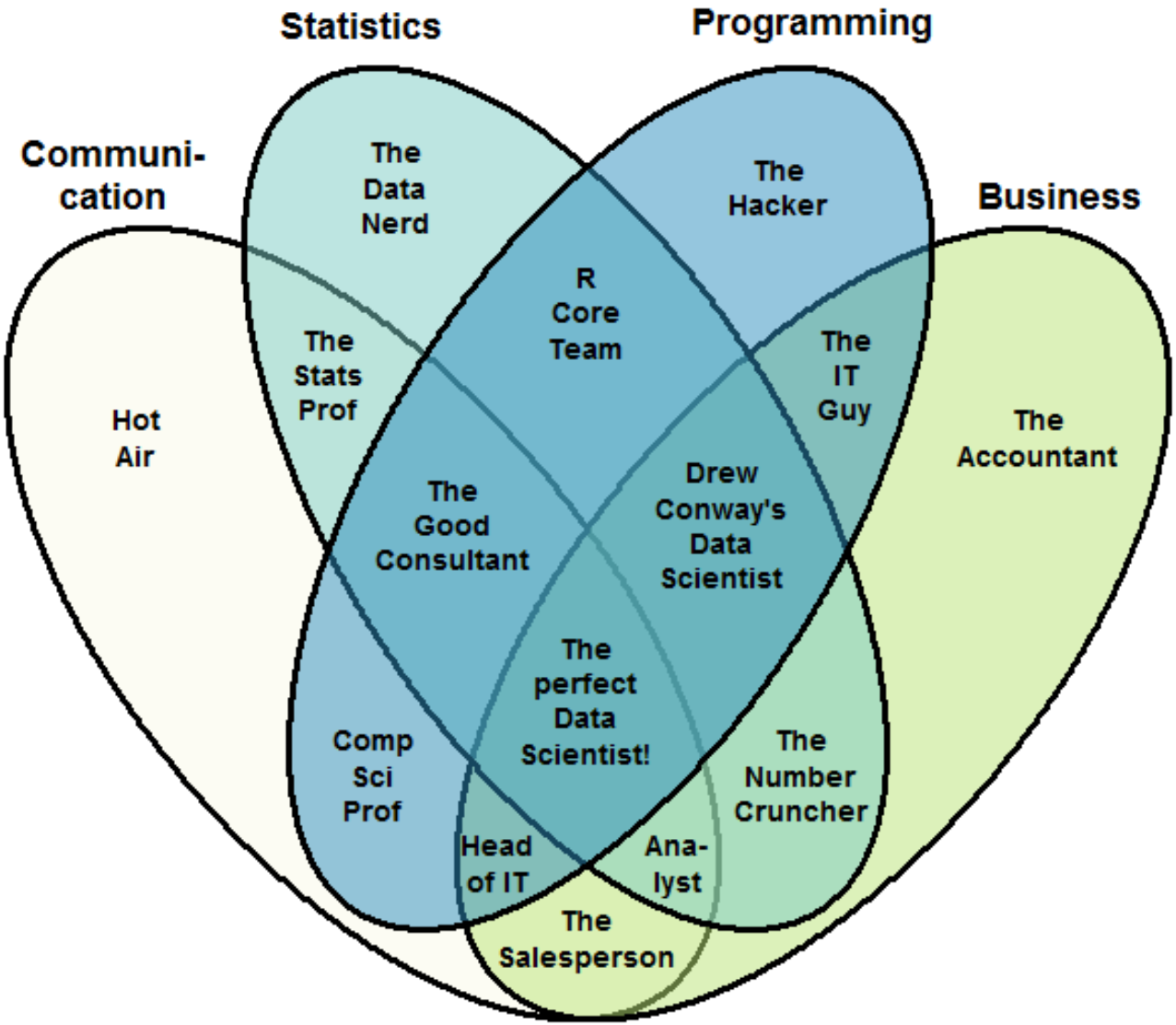
Final Project

Michaël Defferrard Xavier Bresson Pierre Vandergheynst

EPFL LTS2 Laboratory

November 14, 2016

Data Scientist



Project

1. Define a problem.

2. Solve it in a Jupyter notebook using the Data Science process.

3. Handle your solution for grading.

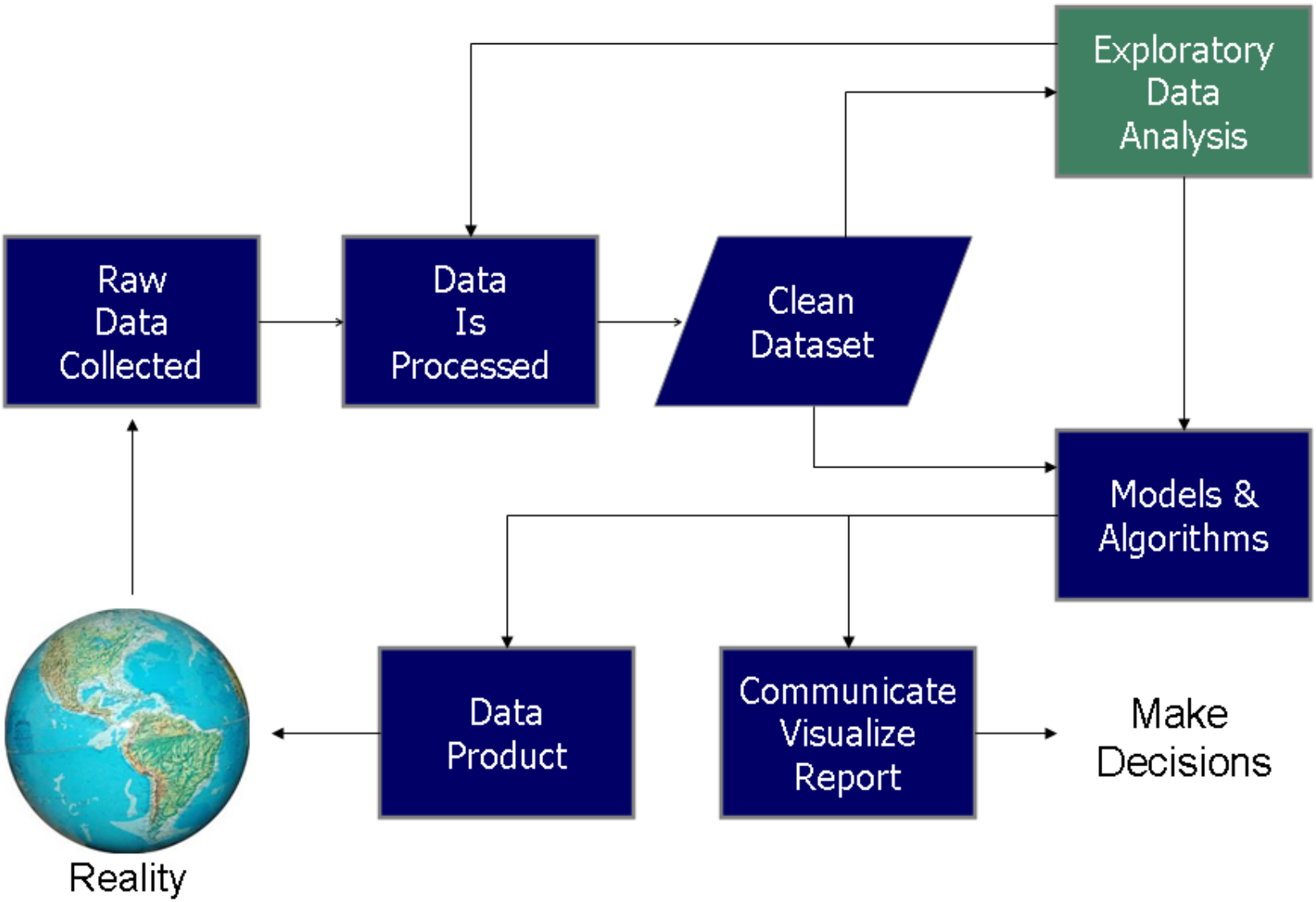
Problem

Find a problem you want to solve.

- Classic problems: Computer Vision, Natural Language Processing.
- Social websites are a wealth of information.[1]
- Your own interests: scientific, hobbies or otherwise.
- Open challenges, e.g. kaggle.
- Any other.
- Discuss with us!

[1] Not only Facebook and Twitter, but also GitHub, Pinterest, StackOverflow, YouTube, LinkedIn, Instagram, Tumblr, etc.

Data Science Process



Structure

The structure of the notebook will follow the Data Science process seen during the exercises.

1. Data acquisition: from the web, a database, a flat file, etc. This includes cleaning the data.
2. Data exploration: simple exploratory analysis to describe what you got.
3. Data exploitation: build and train a Machine Learning algorithm based on this data. Any algorithm is considerable, but it has to be motivated.
4. Evaluation: evaluate the performance, e.g. accuracy, training time, etc., of the chosen model. You define the metrics you care about! If you tried multiple algorithms, please report their performance and try to explain it.

Practical aspects

- Please isolate code blocks in functions and put those in a separate Python module.
- Your notebook should be clean and legible.
- You can take inspirations from the notebooks seen during the exercises.

Organization

1. Project proposal:
   - Single page document explaining what you want to do.
   - Organize yourselves in groups of one, two or three people.
   - Deadline: Sunday, November 27, 2016. Upload on Moodle.
   - Not graded.
2. Solution:
   - Jupyter notebook with text, math, code, analyzes and results.
   - Deadline: Sunday, January 15, 2017. Upload on Moodle.
   - The notebook will be posted on the course git repository, on GitHub. You can use it for your portfolio!
   - Graded.
3. Presentation:
   - Presentation of 15 minutes in front of the class.
   - Date to be announced, after January 15.
   - Graded.