

# FOUNDATIONS OF HEALTH DATA SCIENCE



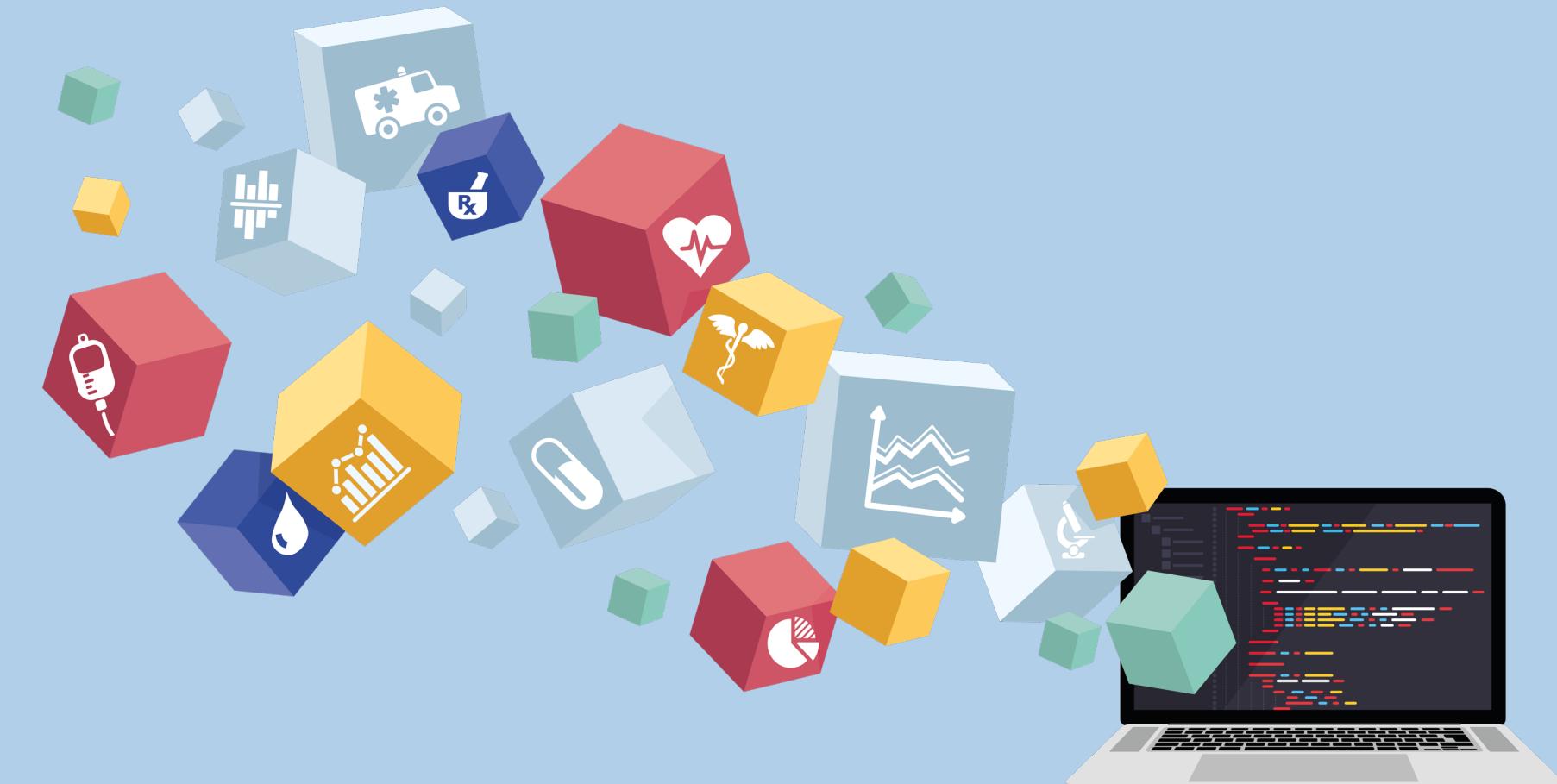
An Introduction for SUND  
Researchers & Educators

# PROGRAM

08:15	MORNING COFFEE
08:30	INTRODUCTION TO HDS
09:15	DATA COLLECTION
10:00	COFFEE BREAK
10:15	EXPLORATORY DATA ANALYSIS
11:30	DATA ANALYSIS PART 1
12:00	LUNCH
13:00	DATA ANALYSIS PART 2
13:30	MODEL EVALUATION PART 1
14:15	COFFEE BREAK
14:30	MODEL EVALUATION PART 2
15:00	WRAP-UP & DISCUSSION

## COURSE MATERIALS:

<https://github.com/Center-for-Health-Data-Science/Foundations-of-Health-Data-Science>



# Center for Health Data Science (HeaDS)

---

The **mission of the Center** is to strengthen **health data science** within the Faculty:

- Active and visible hub for Health Data Science
- Providing data science support for researchers at SUND
- Courses, workshops and training environments to improve data science skills
- Support a network of researchers and educators



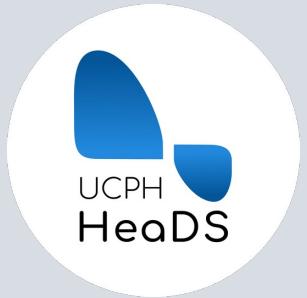
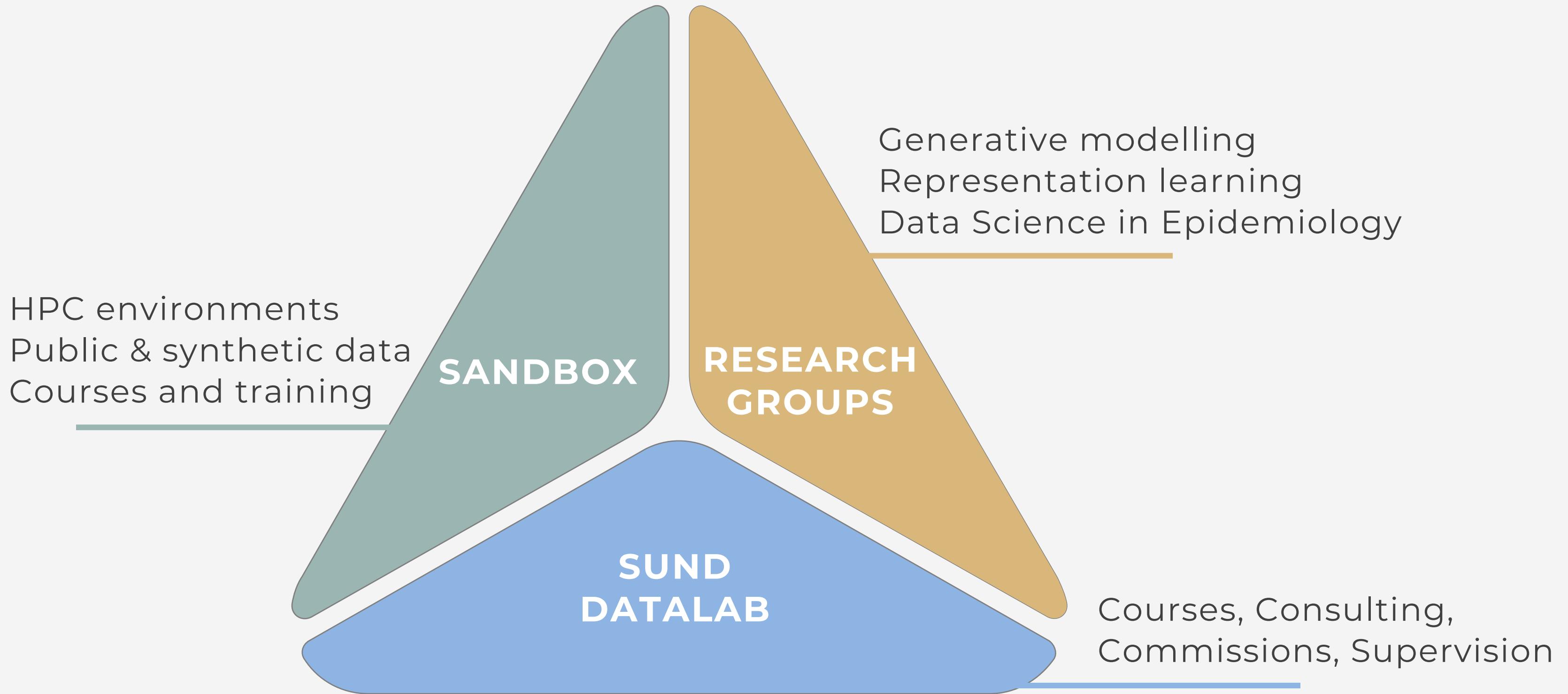
Henrike



Thilde



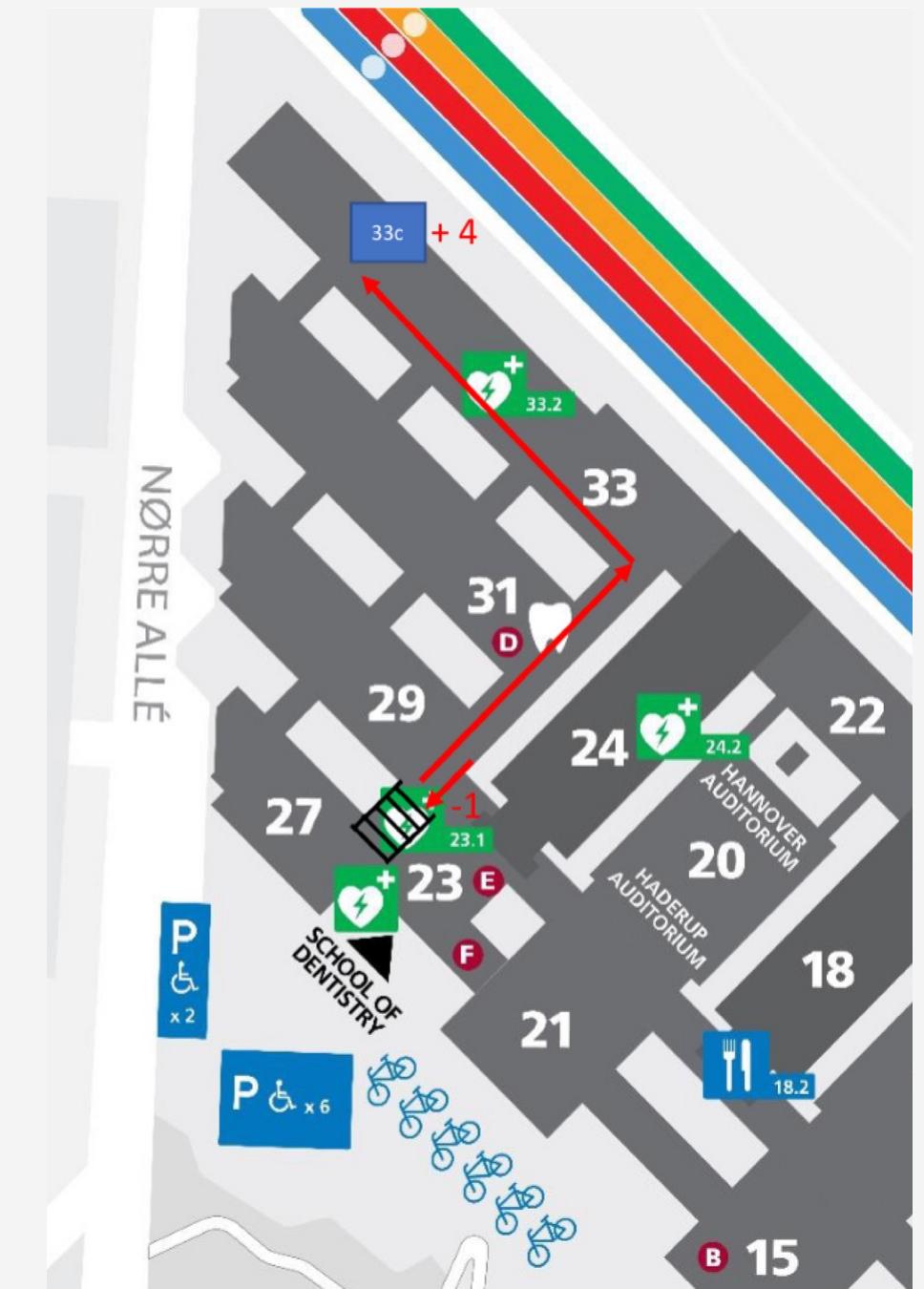
# CENTER STRUCTURE





# CONTACT US

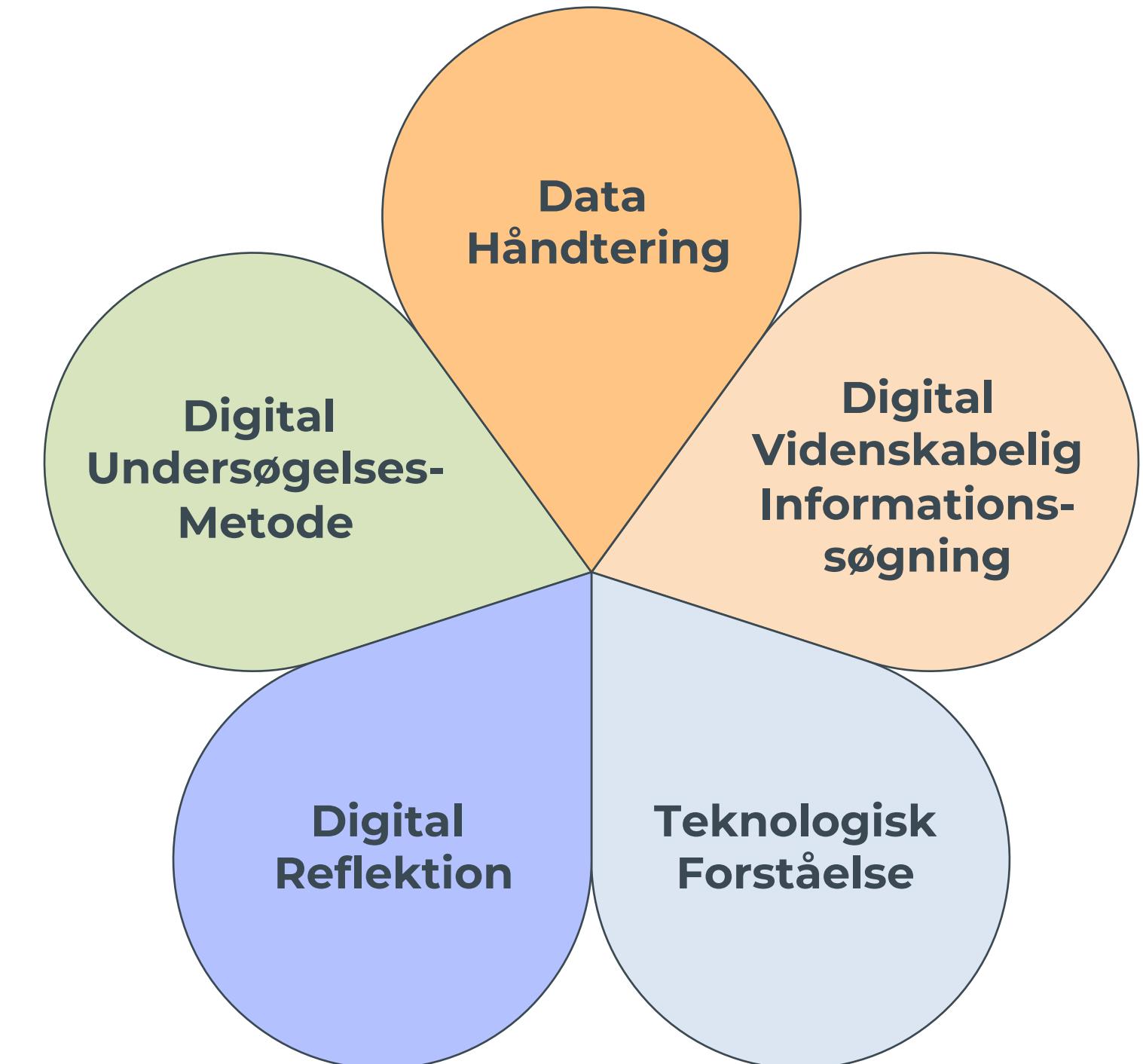
- Email:** heads-admin@sund.ku.dk | datalab@sund.ku.dk
- Website:** <https://heads.ku.dk/>
- Location:** Building 33, 4. floor, Section C, Panum



# DIGITAL CORE CURRICULUM INITIATIVE

The Digital Core Curriculum (DCC) initiative

- KU-wide initiative, started at SUND
- Update all KU educations to include digital literacy skills and data science competences
- Each study board does its own implementation. DCC working group supports and advises
- Goal: Implementation of a Digital core curriculum corresponding to 5 – 7,5 ECTS per study line

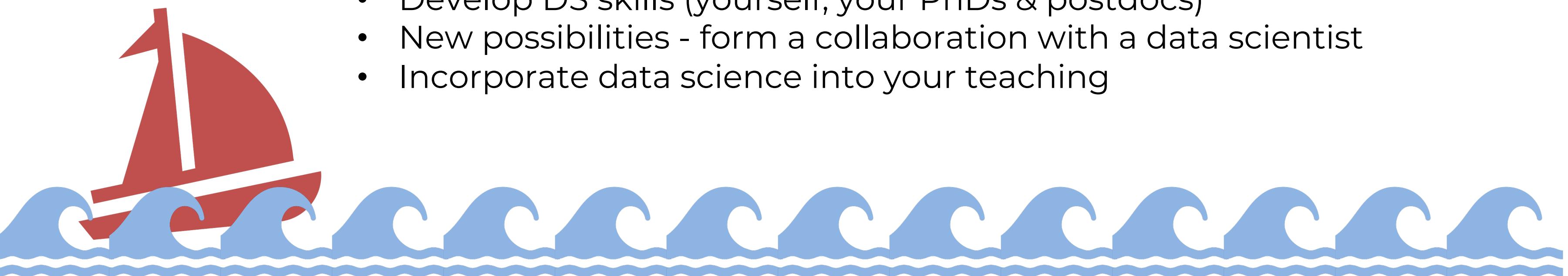


# THE PURPOSE OF THIS COURSE

---

**In this course, we are going to begin our journey into Data Science**

- **What is (Health) Data Science (DS)?** Roles, definitions, data types
- **A Data's journey:** From data collection to scientific results?
- Data science is not scary or mystical. Now you know the 'fancy words'.
- Inspiration:
  - Develop DS skills (yourself, your PhDs & postdocs)
  - New possibilities - form a collaboration with a data scientist
  - Incorporate data science into your teaching

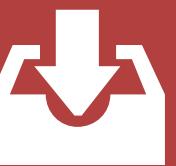


**In this course, we are going to begin our journey into Data Science.**

On our way we shall touch upon data collection, exploration, analysis and evaluation.



**DATA COLLECTION**



**EXPLORATORY DATA ANALYSIS**

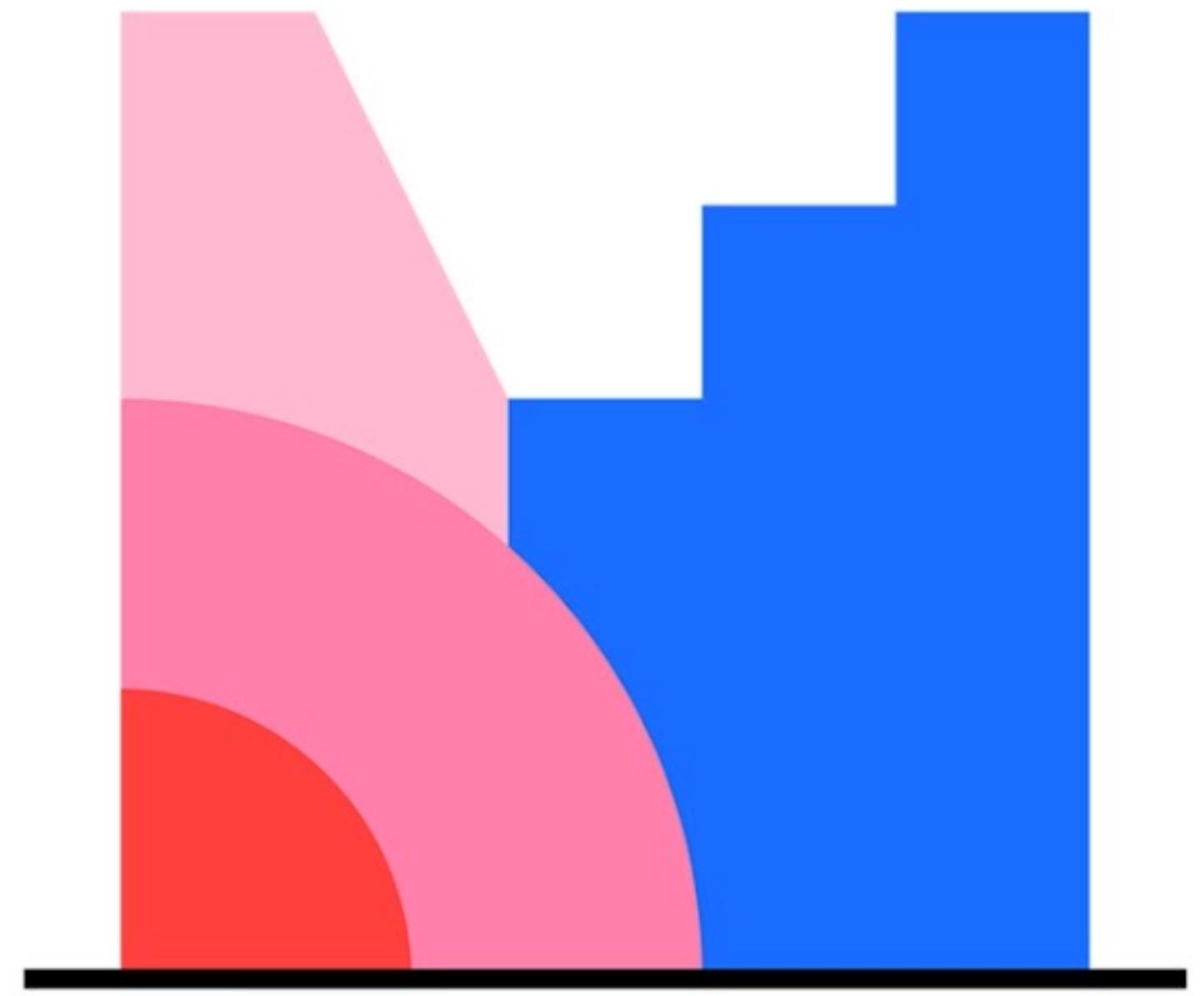


**DATA ANALYSIS**



**MODEL EVALUATION**



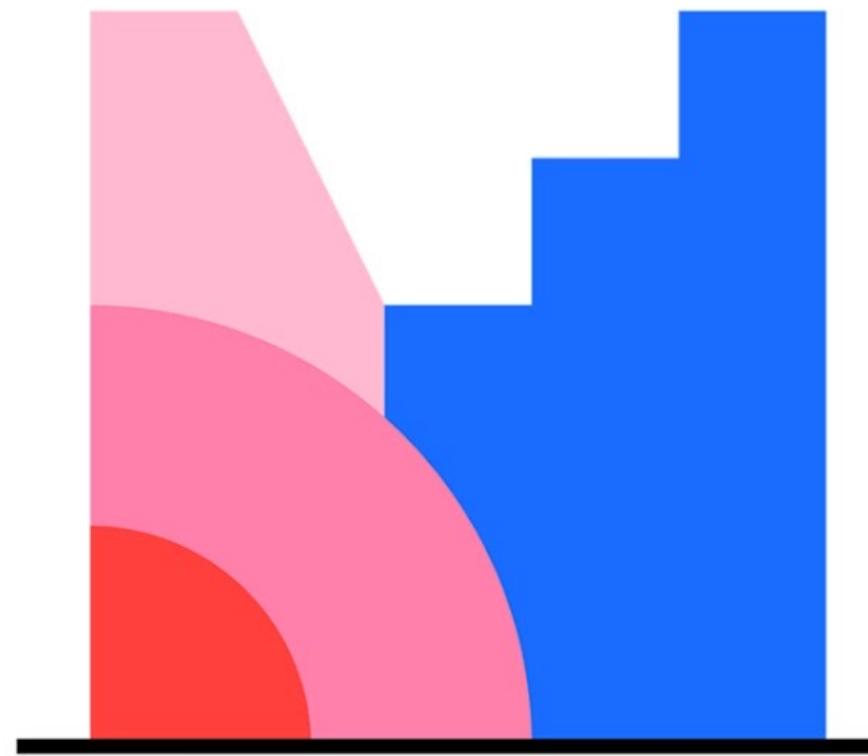


# Mentimeter

During the course we will use **Mentimeter** for feedback and discussion.

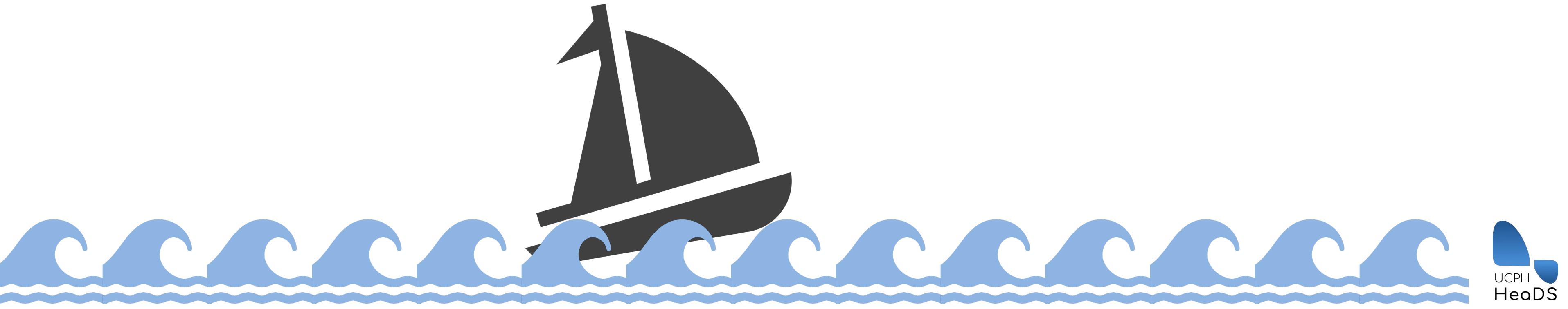
**Let's try it out!**





# Mentimeter

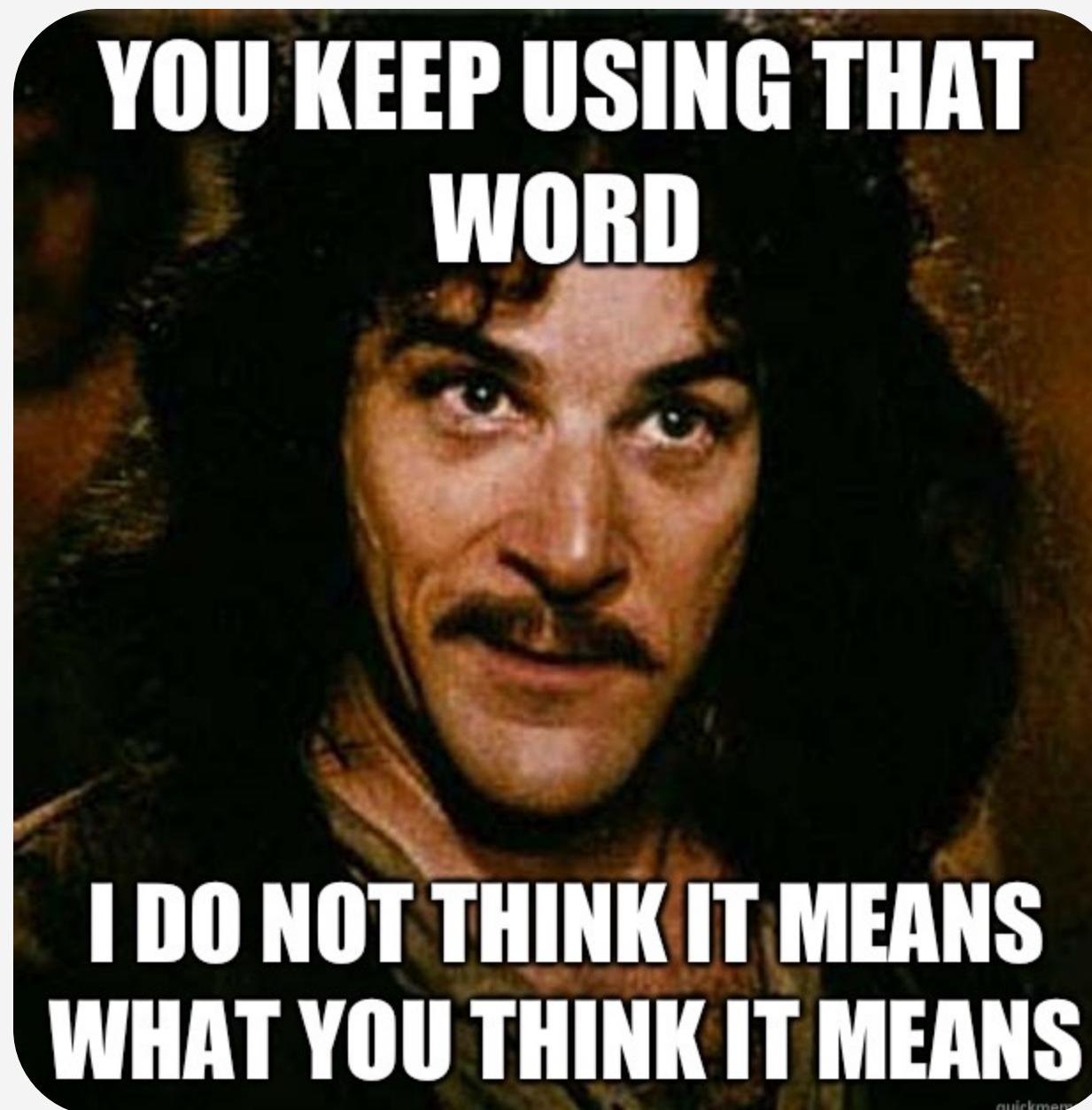
Now that you know about **HeaDS** and this course, we want to hear some things about **you**.



# WHAT IS DATA SCIENCE?



# WHAT DO THE WORDS MEAN?



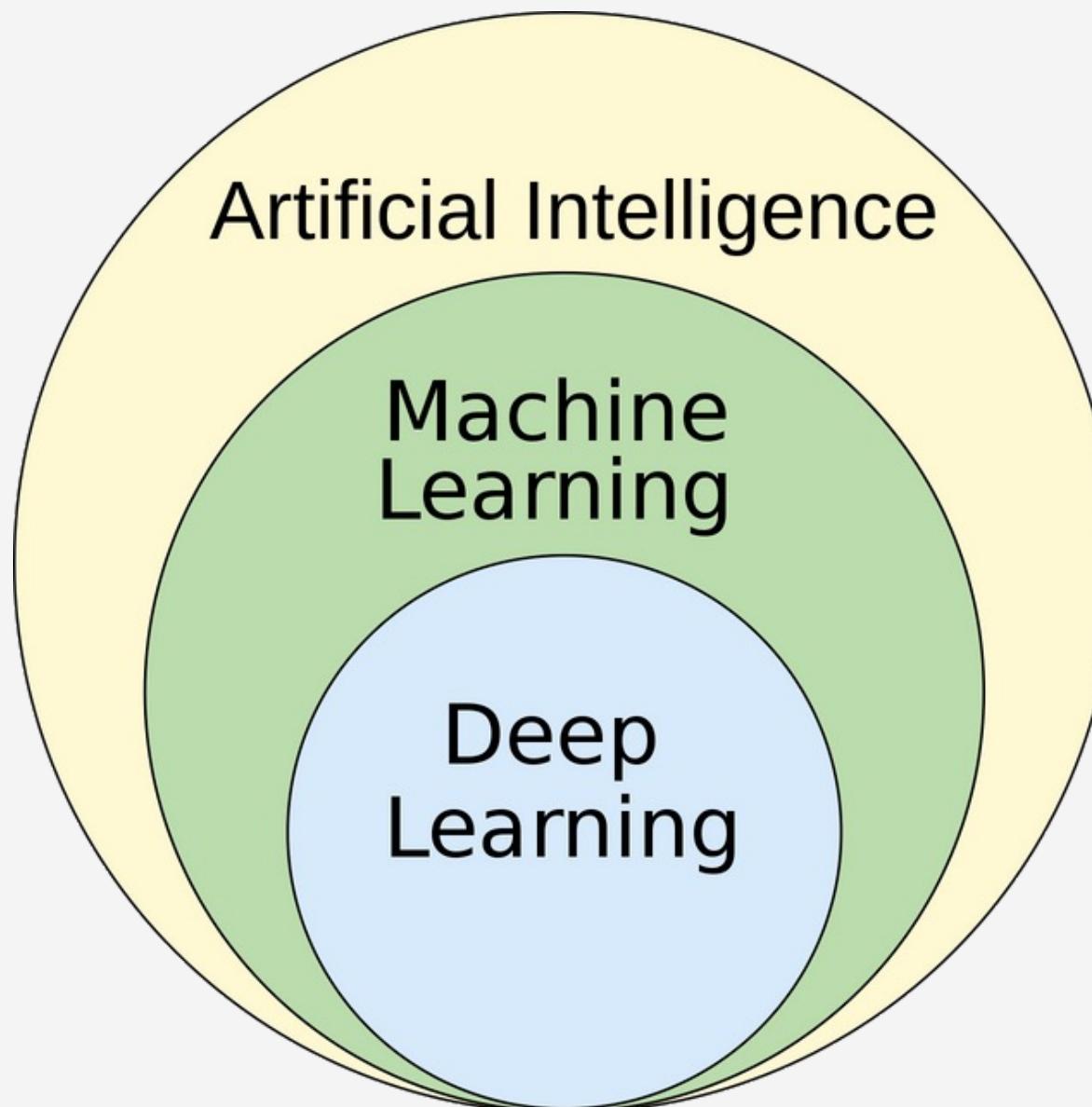
We often hear the words **Data Science**, **Machine Learning** and **AI** used together, often as if they were synonyms.

**Buuuut, they are not :)**

Let's try some differentiation.

# WHAT DO THE WORDS MEAN?

---



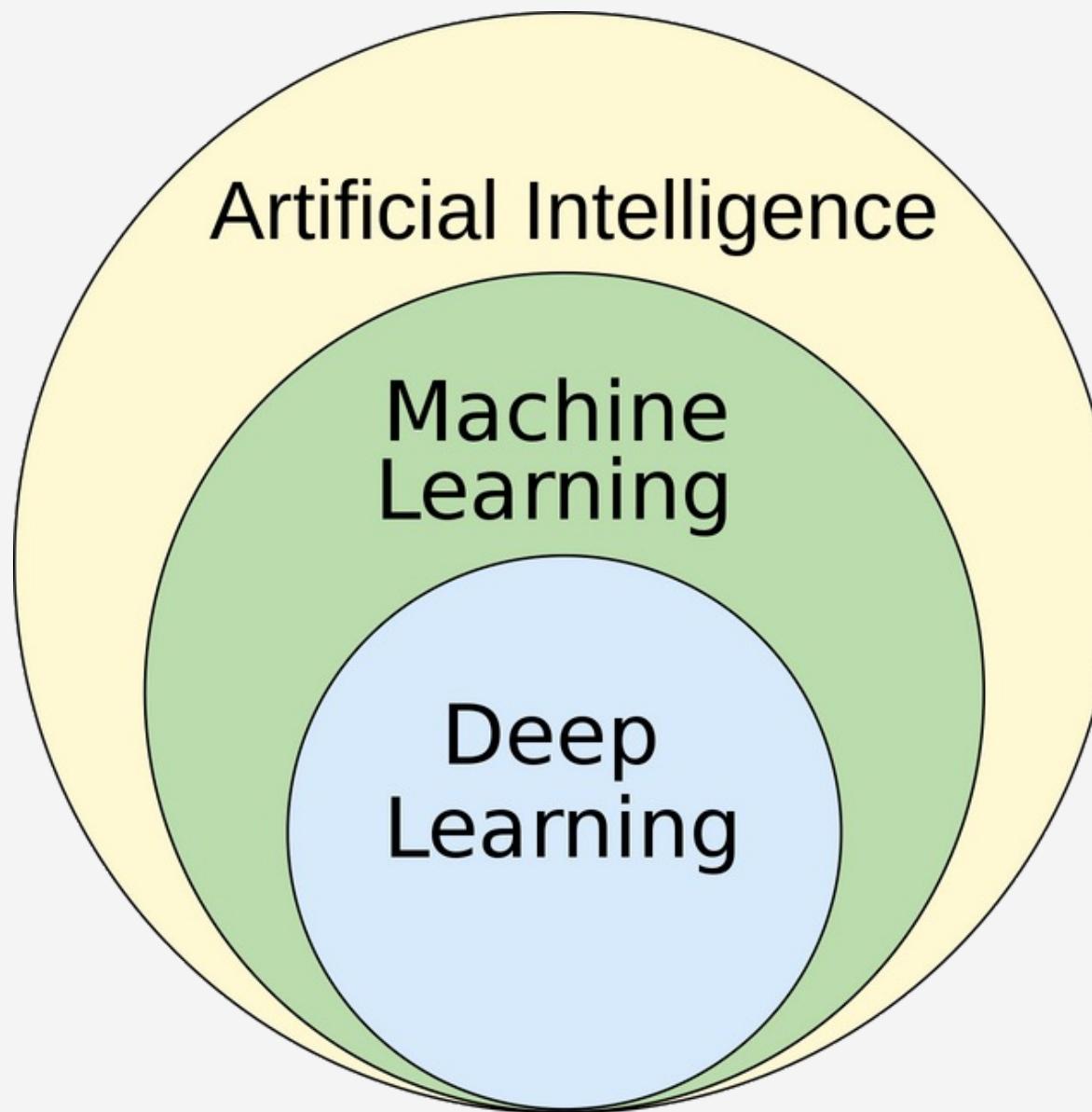
## DATA SCIENCE != MACHINE LEARNING

**Machine learning** - development of algorithms that enable computers to learn from data and make predictions or decisions (with human-like performance or better).

**Data Science (DS)** - a variety of techniques for extracting knowledge from data. This involves cleaning, normalizing, and analyzing large data to uncover patterns and trends.

# WHAT DO THE WORDS MEAN?

---

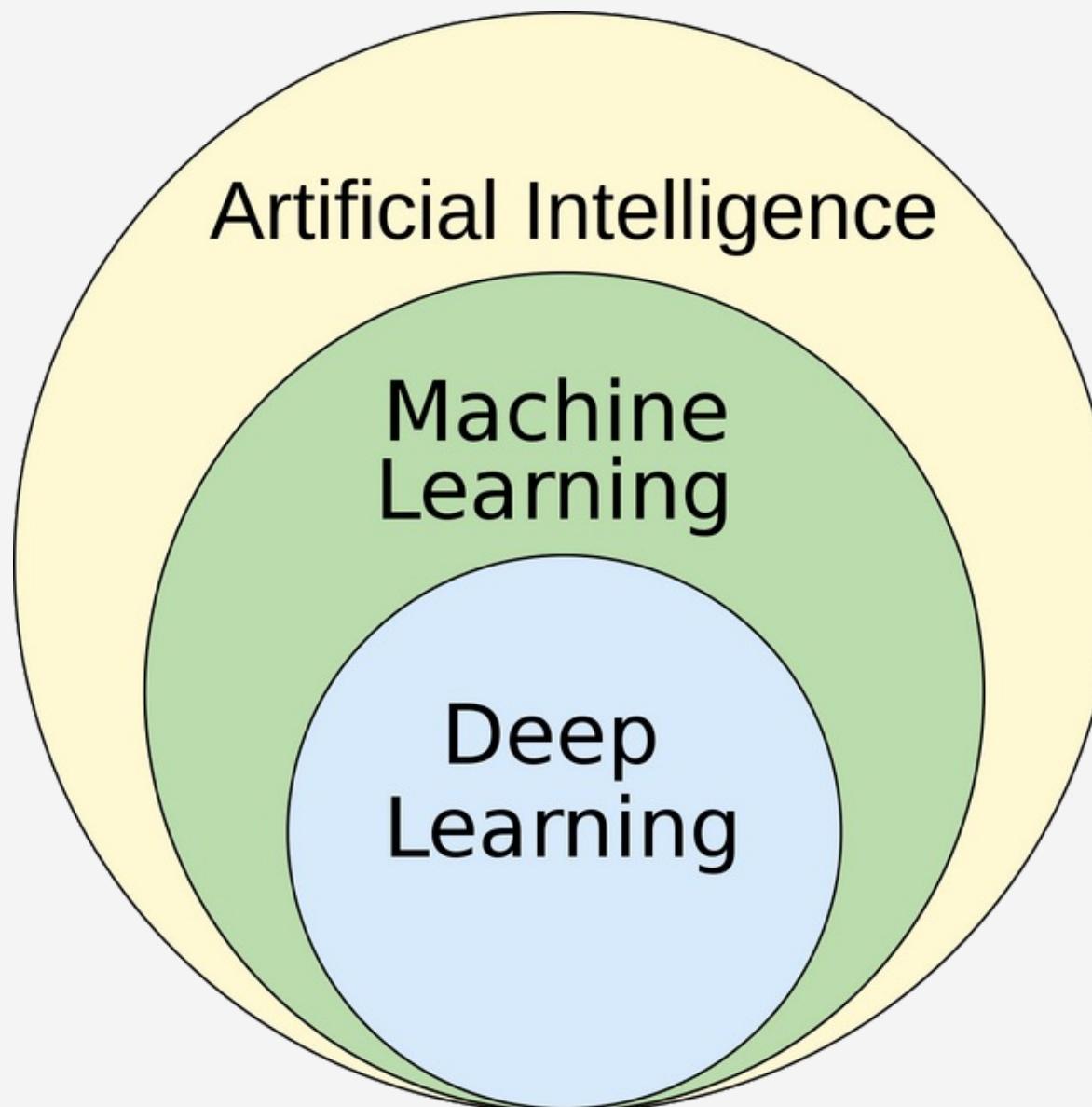


## Artificial Intelligence:

- The ability of computing systems to achieve human-like performance on complex tasks
- Conceptual umbrella term
- AI is the **outcome, not the method**

# WHAT DO THE WORDS MEAN?

---



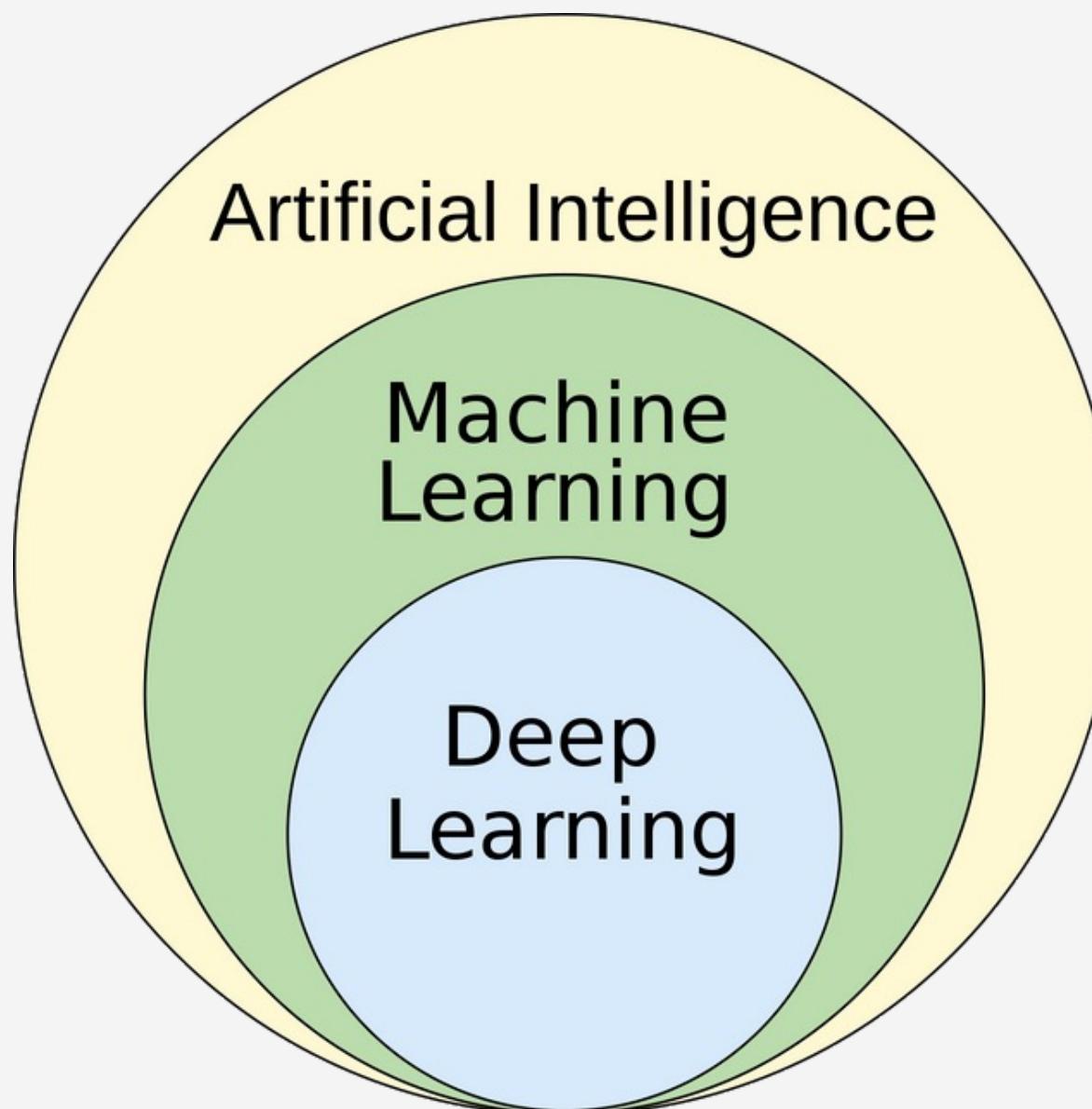
## **Machine Learning:**

- "Technologies and algorithms that enable systems to identify patterns, make decisions, and improve themselves through experience and data" [1]
- Machine learning is methodology.
- Currently our most successful way of achieving AI

1. <https://ai.engineering.columbia.edu/ai-vs-machine-learning/>

# WHAT DO THE WORDS MEAN?

---



## Deep Learning:

- When you do **ML** by means of a **deep neural network**
- This is a type of/sub-class of (ML).

**Data science** is often taken to mean ML/AI, but the term is much broader than that!

...and this is not a course on ML/AI.

# WHAT IS DATA SCIENCE?

**Data Science** combines math, statistics, programming and algorithms with **domain expertise** in order to extract insights from data.

- IBM

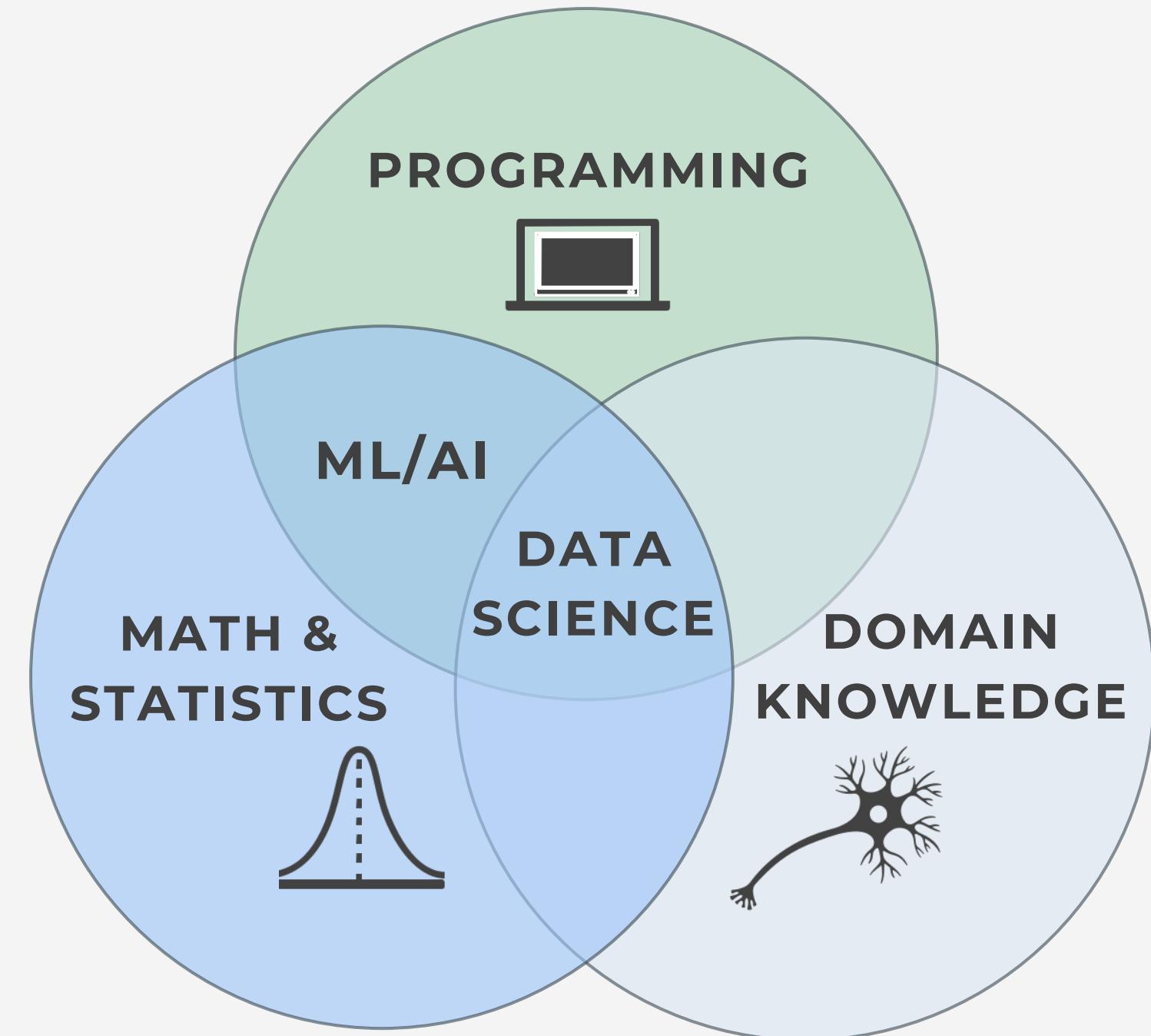
**Data Science** is the processing of data with the goal to **learn something** about its characteristics or answer a scientific question.

- HeaDS



# WHAT IS DATA SCIENCE?

- A **cross-disciplinary** undertaking that draws on many disciplines and is in turn becoming part of many disciplines.
- Data matters, one size analysis does not fit all.



# WHY DO WE WORK WITH DATA?

**Goal:** To extract knowledge about how the world works and, if possible, make generalizations and predictions.

Data Science is the more **formalized process** where we make use of tools such as computers and algorithms to help us make sense of **vast amounts data**.



# ROLES

---

Not everybody is involved in every step.

## Data Collector:

Produces or gives access to the data. Often has domain knowledge on the data, i.e. doctors working at the hospital

## Principal Investigator:

Introduces the research question



## Statistician/Mathematician:

Selects the appropriate tests and/or models  
May do the data science analysis

# ROLES

---

## Data Scientist (Computer Scientist / Bioinformatician)

Does the data wrangling, cleaning and pre-processing

Does the data analysis

May standardize and/or implement as software

## Visualisations expert:

*Makes the plots to illustrate results*

*Selects proper type of plot*

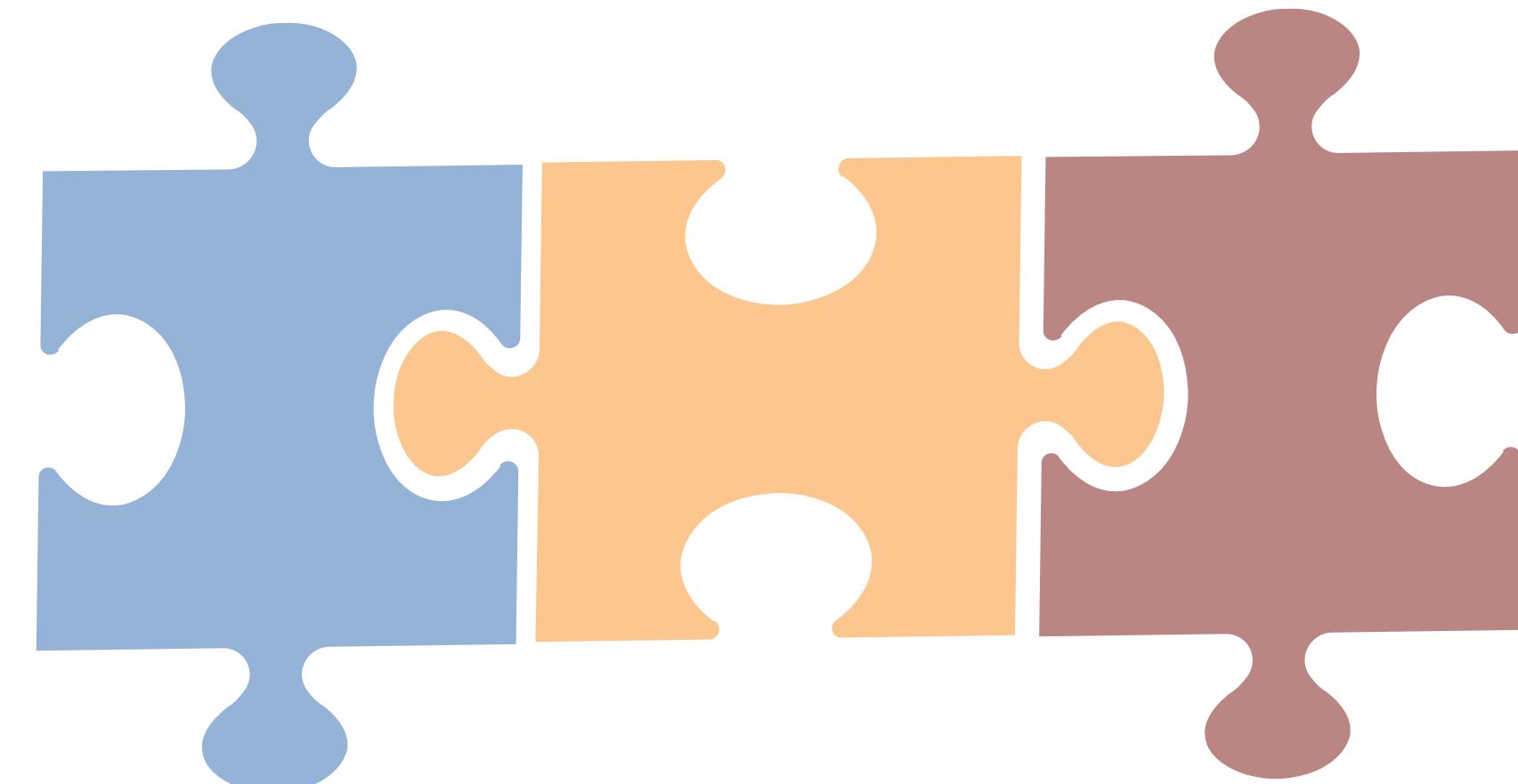
*Focus on interpretability and accessibility*

## Domain expert:

Sparring about results

Biological/clinical relevance

Scientific Publication



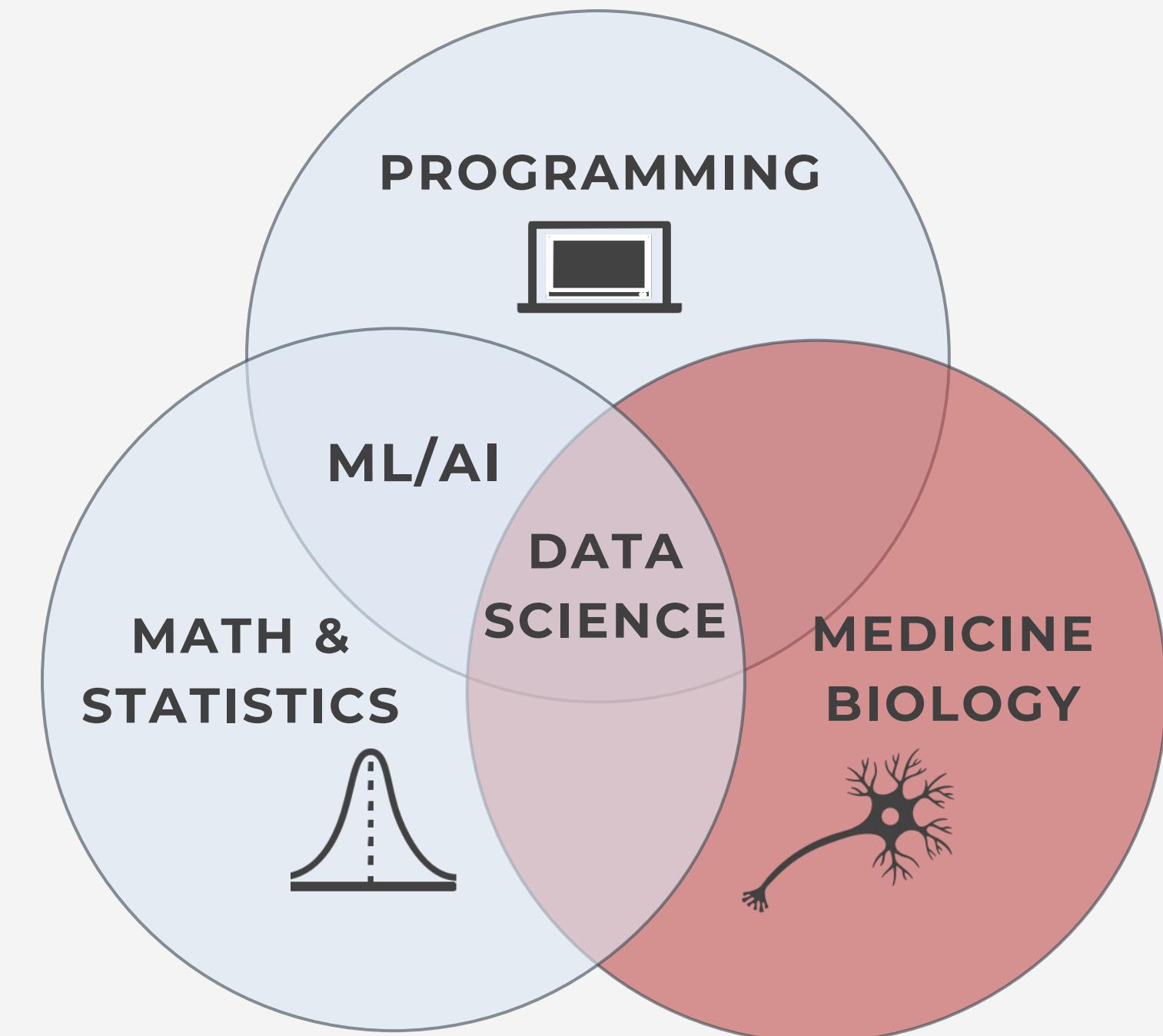
# WHAT IS HEALTH DATA SCIENCE?



In **Health Data Science** the domains of interest are medicine (micro)biology, biochemistry, etc.

## Field is concerned with:

- Biological mechanisms central to disease development
- Discovery and assessment of disease specific drug treatment
- Disease progression and patient survival
- **Personalized medicine**



# WE HAVE THE WHAT – NOW THE WHY?

---

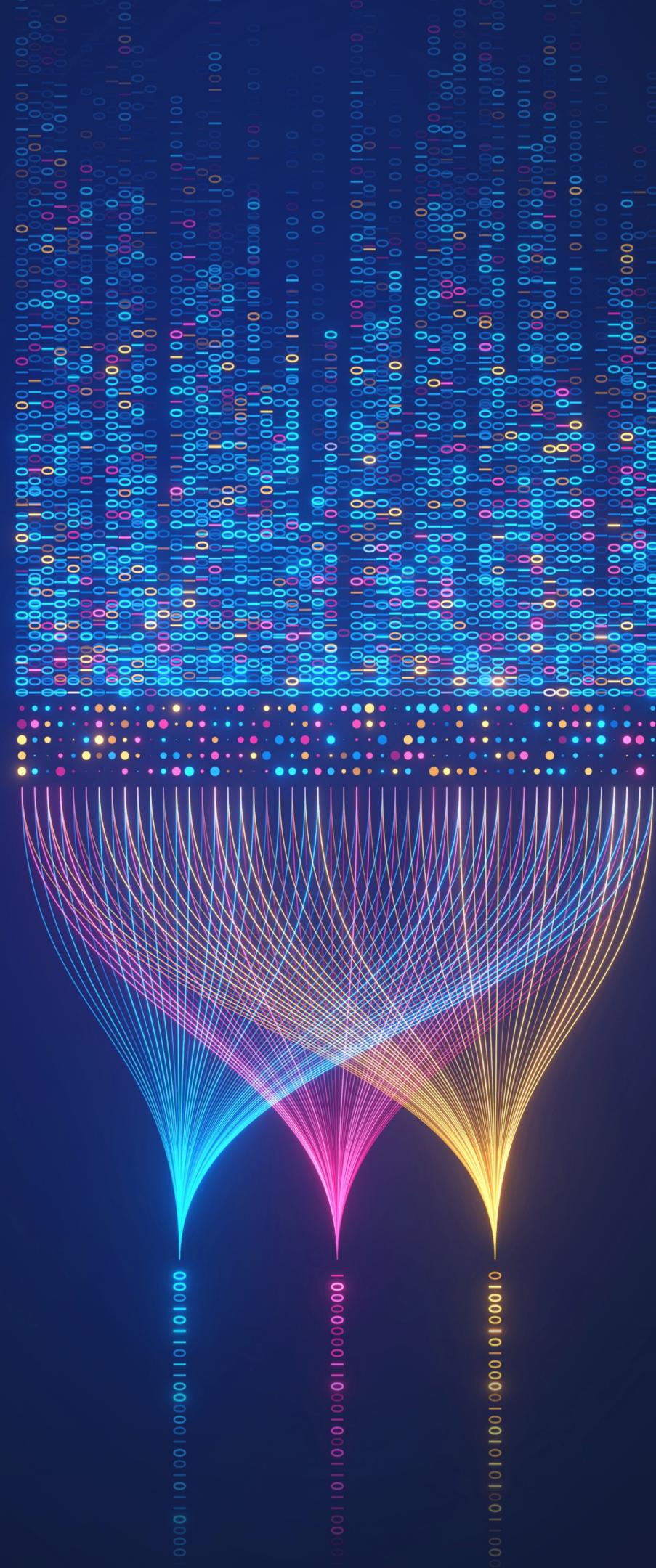
**The world is becoming data driven!**

Amount and quality of data is growing every year

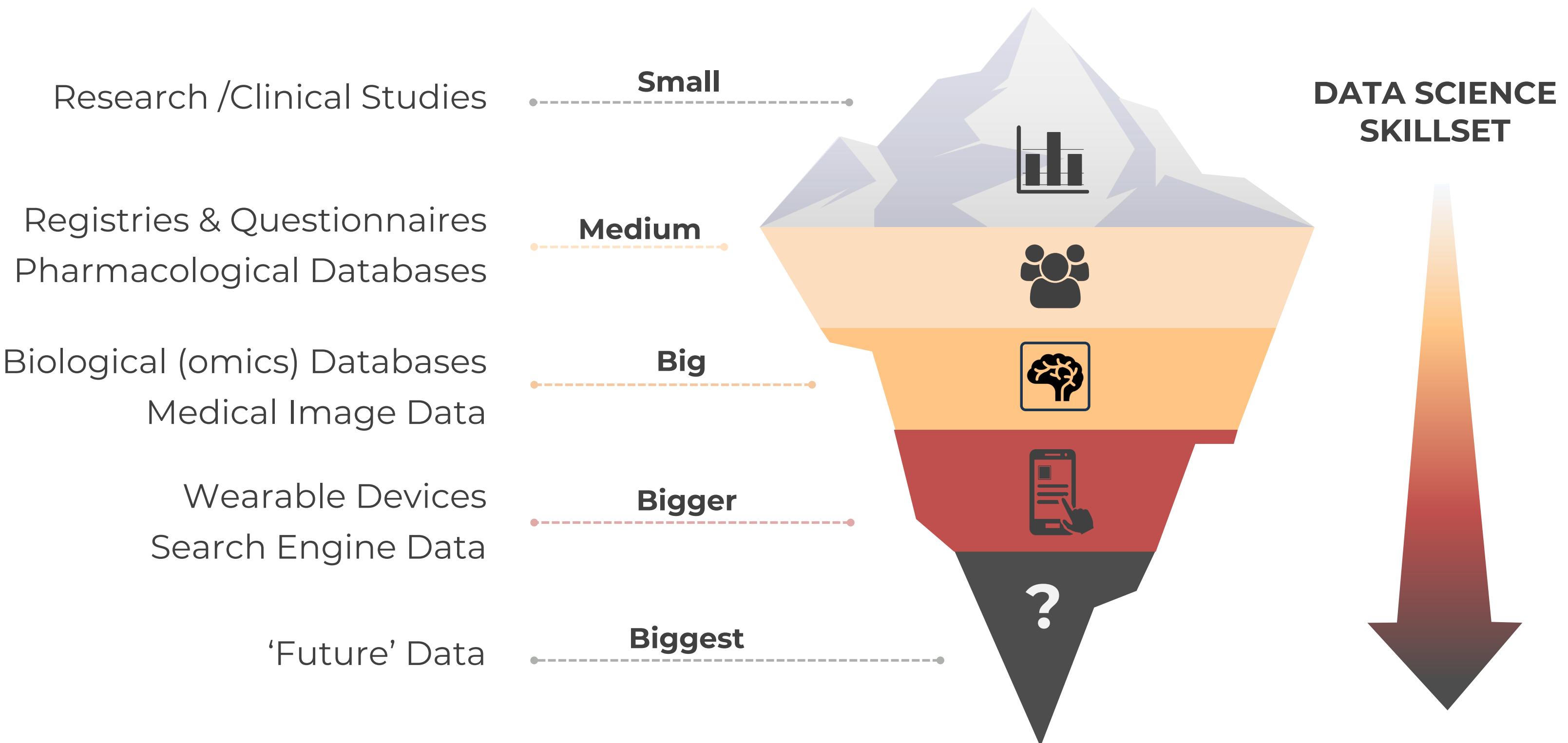
Now data drive research questions & theory

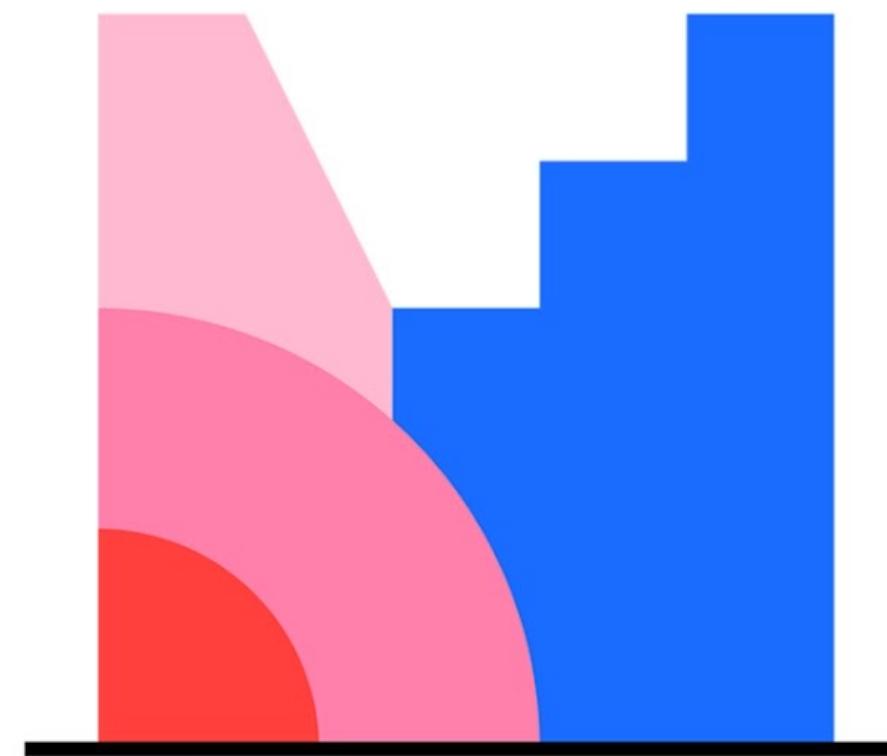
**Improve your science**, as well as your **CV**

**Teach the next generation** how to utilize big data!



# ANALYSIS OF BIG BIO-MEDICAL DATA

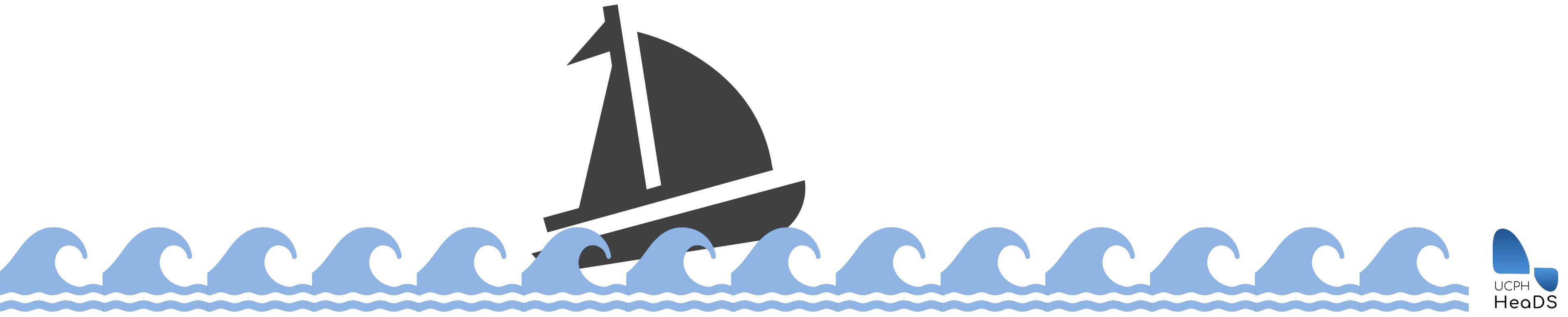




---

# Mentimeter

After this introduction, how do you see the importance of data science to you personally?



# GROUP DISCUSSION

Which of the **roles** we have introduced **do you see yourself in?**

Do you have people in your group or among your collaborators to fill the other roles? If not, **what are the alternatives?**

