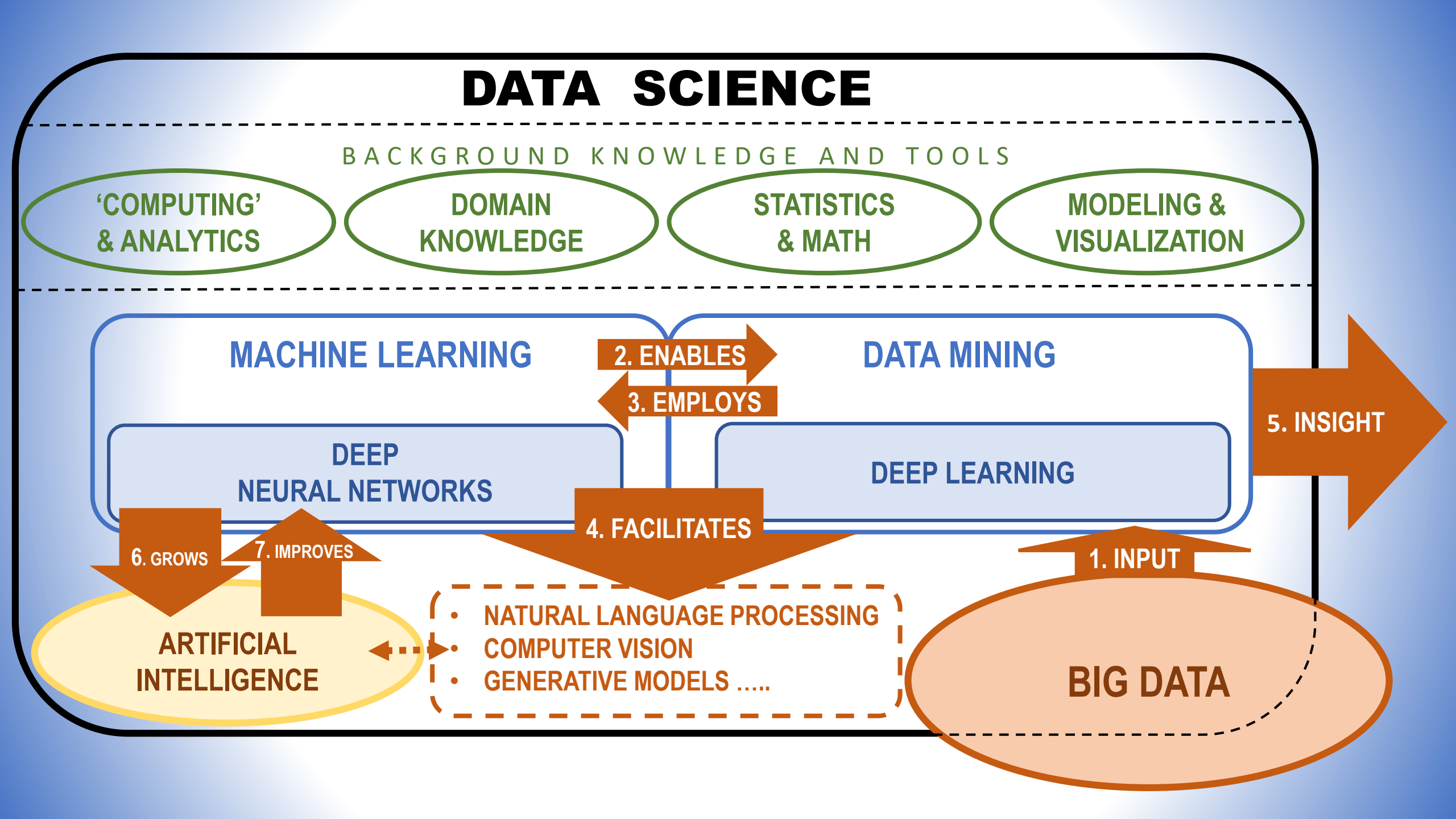
DATA SCIENCE AND ITS TERMS

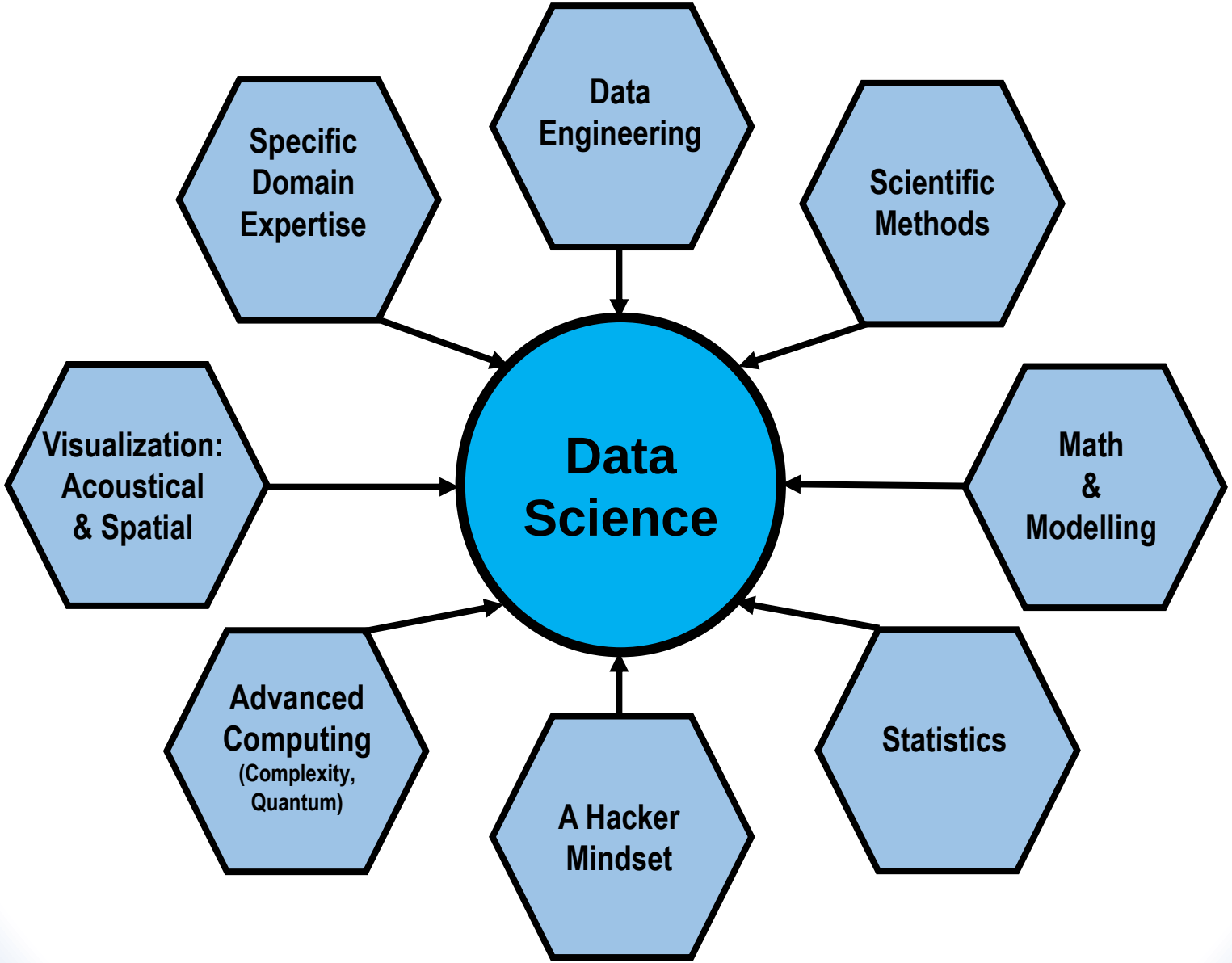
Richard Bruno
CIRANO
v1.4

Definition : Data Science

- Covers **the whole spectrum of data processing**
- It uses a combination of statistics, math, programming, problem solving and capturing data in ingenious ways
- It uses these to cleanse, prepare, & align data for specified needs.
- **Note**: today, in the OECD each person creates 4TB (4*10**12)/mo


DATA SCIENCE
BACKGROUND KNOWLEDGE AND TOOLS
'COMPUTING' & ANALYTICS
DOMAIN KNOWLEDGE
STATISTICS & MATH
MODELING & VISUALIZATION
MACHINE LEARNING
DATA MINING
2. ENABLES
3. EMPLOYS
DEEP NEURAL NETWORKS
DEEP LEARNING
5. INSIGHT
4. FACILITATES
6. GROWS
7. IMPROVES
1. INPUT
ARTIFICIAL INTELLIGENCE
NATURAL LANGUAGE PROCESSING
COMPUTER VISION
GENERATIVE MODELS .....
BIG DATA

Its facets are:



Definition of a Key Area: Big Data

- Big Data:
  - Refers to **handling of very large unstructured** (*& structured*) data sets in automated ways within acceptable lapse-times.
  - This usually includes data sets with sizes beyond the ability of commonly used software tools (*to capture, curate, manage, & process*).
  - Today this is 100TB-100PB.
    - PC: max today 10TB (*10*10**12*)
    - Human Brain: 5PB (*5*10**15*)
    - A Google: 10**100

Definition of a Key Area: Machine Learning & AI

- Machine Learning:
  - Use of **a set of algorithms that train on large data sets** (*Big Data*) to gain insights that result in **predictions or actions** to optimize some set of problems or systems.
- Artificial Intelligence ("AI")
  - A subfield of data science concerned with **solving tasks that are easy for humans, but hard for computers**.
    - E.g. planning, moving around in the world, recognizing objects and sounds, speaking, translating, performing social or business transactions, doing creative work etc.

Where are They Used and Create Value Today?

- Data Science Algorithms e.g.:
  - Internet searches:

    make searches less user/time intensive, more relevant results
  - Recommendations:

    less footwork, greater correlation with 'your crowd', greater personal relevance
  - Digital advertisements:

    better targeting, stronger (click-*to*-click/through-*to*-buy) decision outcomes, more credible stats & use needs

Where are They Used and Create Value Today?

- Big Data e.g.:
  - Financial services:

    'ist' vs 'sein': the 'dynamics' of more informed & satisfied customers, increased organizational productivity
  - Retail:

    'ist' vs 'sein': the 'dynamics' of understanding and delivering the right-goods to the right-customers at the right-time
  - Communication

    'ist' vs 'sein': the 'dynamics' of getting 'thinking-evaluating-deciding' to become real-time

Where are They Used and Create Value Today?

- Data analytics e.g.:
  - Healthcare

  the factual basis and organization of better outcomes
  - Travel

  efficient organization of complex trips & cost effective disintermediation
  - Gaming

  retaining passion during real-time Massive Multiplayer Online Games
  - Energy management

  real-time serialization and sustainability evaluation and implementation