

# The 'Bachelorisation' of Data Science: What's the Role of EuADS for Standard Curricula?

---

**José Hernández-Orallo**

Escola Tècnica Superior d'Enginyeria Informàtica  
Universitat Politècnica de València (UPV), Spain  
jorallo@dsic.upv.es

**European Data Science Conference  
Luxembourg, November 7-8, 2016**

# How is academia reacting to DS?

---

## ■ Postgraduate degrees (master)

- Academia has reacted relatively fast and it may be aligning with the demand.
  - From dozens to thousands all around the world in the last 5 years.
- However,
  - Many programmes are focused on **very specific** technologies and tools (R, Python, MapReduce, Hadoop, ...) **rather than more general skills** or new soft skills.
  - Master degrees are fine as long-life learning, but **not very efficient** for fresh students coming into the university.
    - Most of the skills and knowledge acquired before the master degrees are not really needed.

# How is academia reacting to DS?

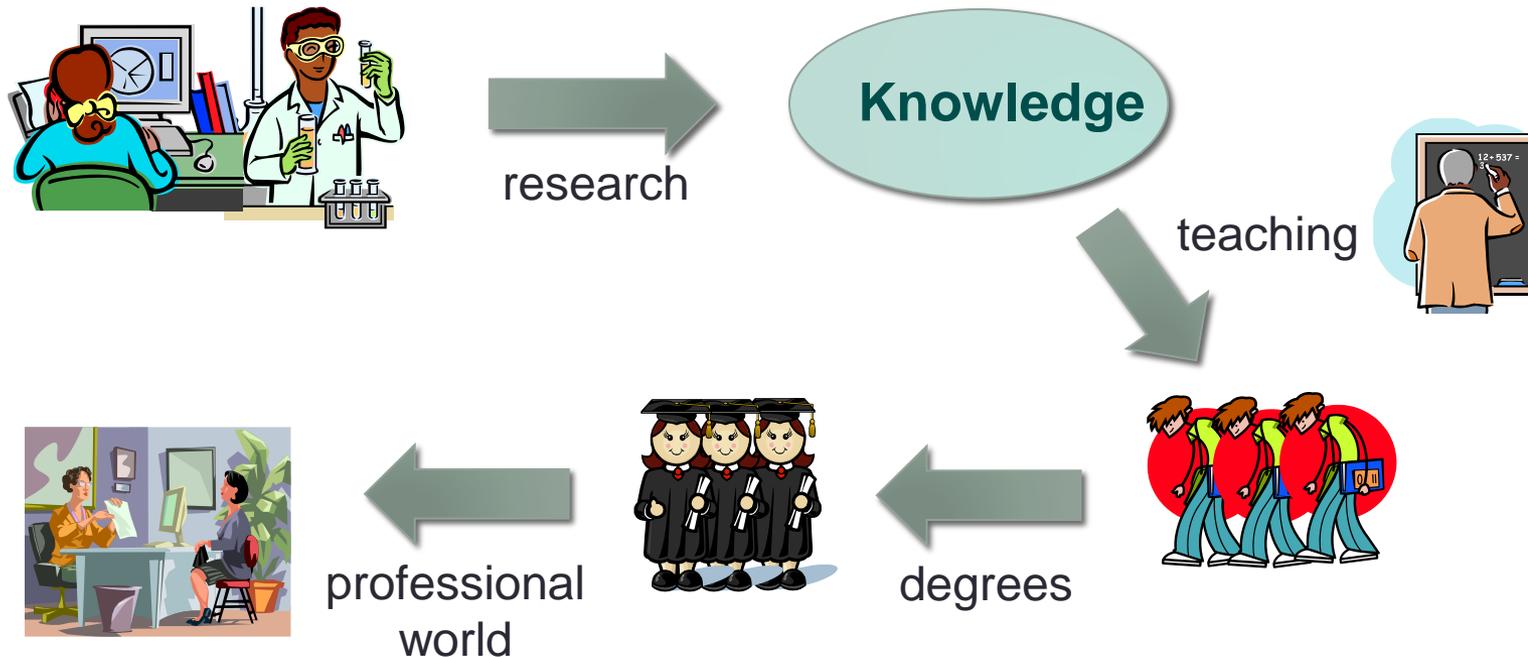
- Undergraduate degrees (bachelors)
  - Academia is now starting to cover this niche.
    - In Europe: from two (Maastricht, Warwick) to dozens in the last 5 years.
    - Adapting statistics and operational research degrees.
  - This “bachelorisation” has not happened with other related areas in the past (data mining, machine learning, ...). Not even “big data”.
    - They weren’t considered a profession.

Data Scientist is proposed as a profession!

- Professions and bachelor degrees are usually associated.

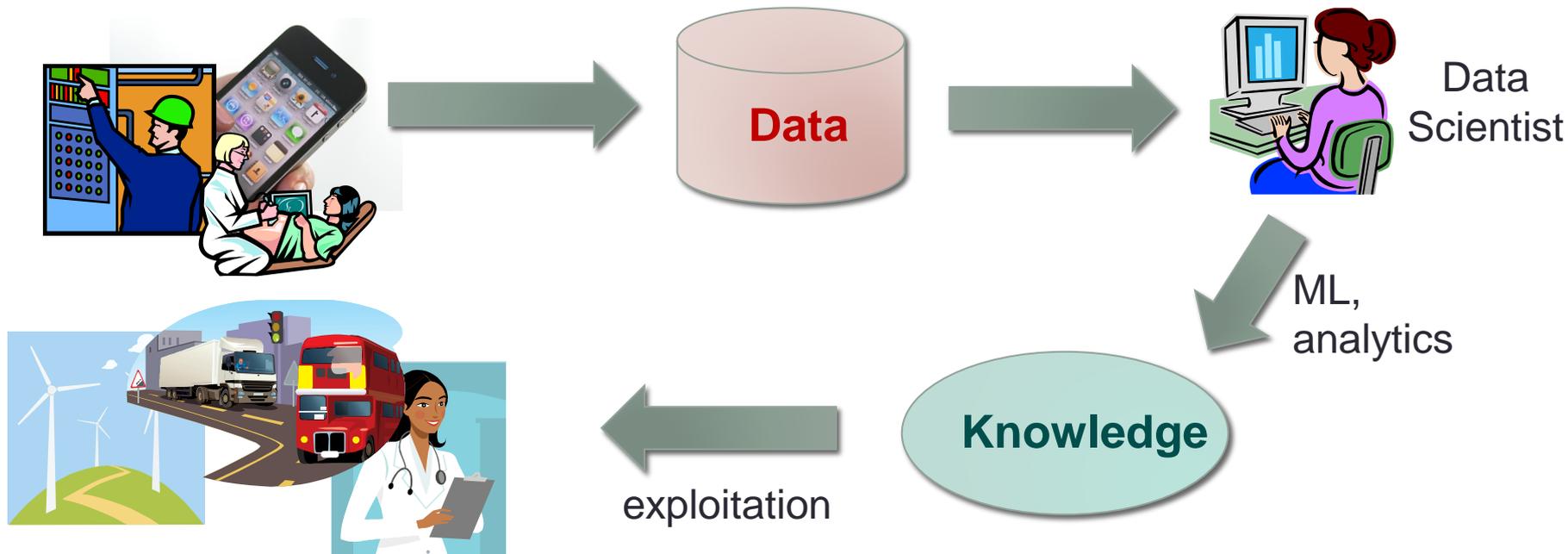
# Is Data Scientist just another (IT) profession?

- Traditional view of the generation of knowledge:
  - Universities generate knowledge through research. Students acquire this knowledge at university. They apply this knowledge as professionals.



# Is Data Scientist just another (IT) profession?

- The future (or present) of the generation of knowledge:
  - People, companies and organisations deal with changing phenomena. Lots of data are stored. New, domain-specific actionable knowledge has to be extracted and deployed.



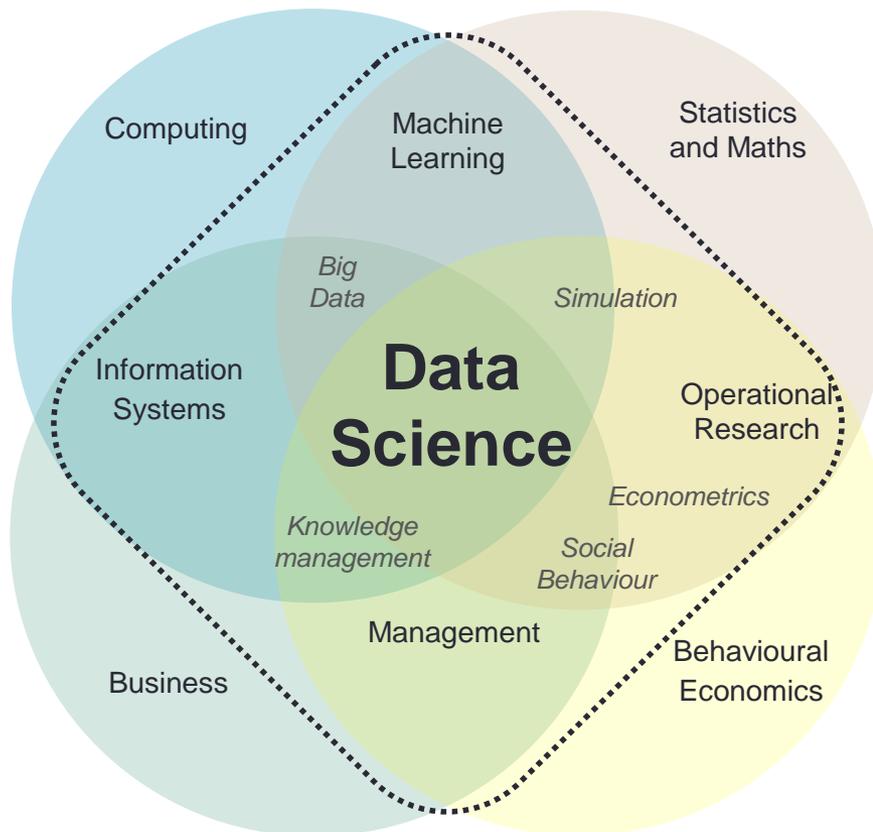
# How must academia **anticipate** this change?

- Under this perspective, a bachelor in data science can be conceived as a threat or an opportunity for universities.
  - Opportunity: Bachelor in data science. If it's well designed!
    - Multi-disciplinary:
      - “spanning mathematics, machine learning, artificial intelligence, statistics, databases, and optimization, along with a deep understanding of the craft of problem formulation to engineer effective solutions”\*.
    - Multi-domain:
      - Aware of the socio-economical context in areas such as medicine, energy, environment, finance, transportation, etc.
    - Multi-empathic:
      - Professionals will need to interact with humans, cognitive systems, robotic systems, hybrids, ...
      - Interaction, presentation and communication are changing.

\* Communications of the ACM “Data Science and Prediction”, December 2013, <http://cacm.acm.org/magazines/2013/12/169933-data-science-and-prediction/fulltext>

# An example at UPV

- Several MSc in data analytics, big data, etc.
- Now designing a 4-year bachelor degree:



- Strong focus on soft skills:
  - insight, curiosity, communication, teamwork, problem solving, creativity and resilience skills,
  - projects every year

# The (draft) syllabus for Data Science at UPV

Semester 1A	Semester 1B
Business Organisation	Linear Algebra
Discrete Mathematics	Statistical models for decision making I
Mathematical Analysis	Programming
Programming Fundamentals	Computers and operating systems
Exploratory Data Analysis	PROJECT: "Data understanding"

Semester 2A	Semester 2B
Statistical models for decision making II	Data management
Economic and social behaviour	Machine learning: Descriptive models
Data structures	Discrete modelling and information theory
Databases	Data security
Data acquisition and transmission	PROJECT: "Data integration and preparation"

Semester 3A	Semester 3B
Digital Economy	Algorithms
Knowledge representation and reasoning	Model evaluation, deployment and monitoring
Machine learning: Predictive models	Natural language and information retrieval
Infrastructure	Continuous modelling
Visualisation	PROJECT: "Data analysis"

Semester 4A	Semester 4B
Machine learning on a large scale	Ethics and the professional and legal setting
Optimisation	Project management
Optional 1	PROJECT: Final Degree Project
Optional 2	Optional 5
Optional 3	Optional 6
Optional 4	

- Optionals arranged in two blocks:
  - Domains: healthcare, biology, education, finance, IoT, Open Data, Retailing, Social Networks, Big Science, ...
  - Advanced courses.

# Discussion

---

- Setting up a new bachelor degree involves a trade-off of different views for a profession, not yet clearly delineated.
  - Discussions (battles) about what a data scientist should be and how to achieve it are sometimes really heated!
- Standard curricula?
  - Would be very useful in the process of setting up a new degree.
  - Would help harmonising data science degrees.
  - Would be key to consolidate data science as a profession.
- Questions for EuADS?
  - Should EuADS think of publishing reference curricula?
  - Is EuADS thinking of accreditation labels for bachelor degrees?