

# Training Data Scientists : a few challenges

Gilbert Saporta, Conservatoire national des arts et métiers, Paris

# Impacts on curricula

- Necessary update
- More information technologies
  - Distributed and parallel computing
- More Machine Learning
- Less big software, more free environments:
  - R, Python, ScikitLearn, Spark etc.
- Related fields
  - Econometrics (cf H.Varian, 2014)
  - Official Statistics (EMOS)

# Shortage of talents

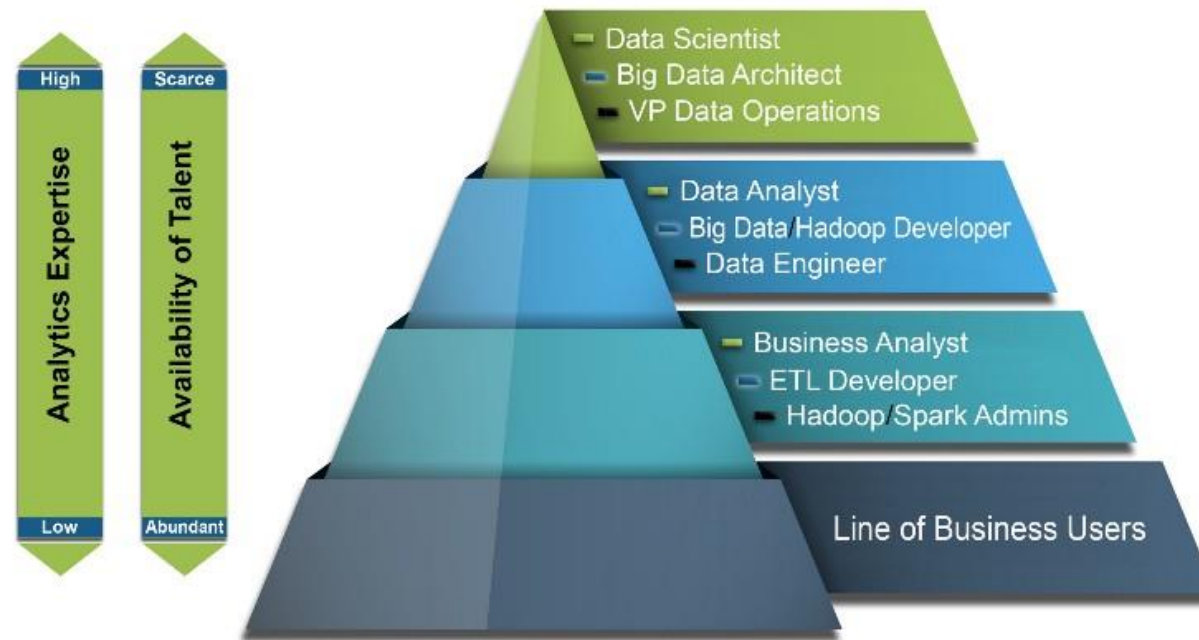
- Kdnuggets (jan.2016)

**Businesses Will Need One Million Data Scientists by 2018**

- There will be a shortage of talent necessary for organizations to take advantage of big data. **By 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills** as well as 1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions. (**McKinsey**)

- **International Data Corporation (IDC)** predicts a need by 2018 for **181,000 people with deep analytical skills**, and a requirement five times that number for jobs with the need for data management and interpretation skills.

### Addressing the Big Data Skills Gap



Roles: **Data Analysis** - **Development** - **Operations**

- According to her estimates, there were about **6.1 million data workers** in the EU, which represents 3% of total EU employment, most being high profile professionals, managers in the field of data innovation. She then revealed that there was a **skill gap of 7.5 percent** of total demand, in other words vacancies in 2014, between the supply of data workers in Europe and the demand for data workers, and that that represents a threat to the European data industry. **It became apparent that the most relevant gap concerned data scientist skills.** “Data scientists” were defined as people with special technical, mathematical tools specialised for designing tools and applications for data and there are about 100 000 -150 000 in Europe of these highly specialised scientists.

Gabriella Cattaneo, IDC European Government Consulting  
European Data Forum 2015, Luxembourg, 17 November 2015

- over the past 5 years the demand for Big Data staff in the UK alone had risen tenfold and currently 77% of Big Data roles are hard to fill. It is also estimated that there will be a 160% increase in demand for Big Data specialists between **2013-2020 to 346,000 new jobs.**

( <https://ec.europa.eu/digital-single-market/events/cf/ict2015/item-display.cfm?id=15762>)

The screenshot shows the website for the Digital Single Market, specifically the Digital Economy & Society section. The header includes the European Commission logo and the text 'DIGITAL SINGLE MARKET Digital Economy & Society'. A navigation menu at the top lists 'The strategy', 'Economy', 'Society', 'Access & connectivity', 'Research & innovation', and 'DG CONNECT'. The main content area is titled 'DG CONNECT' and features a sidebar with links for 'Who we are', 'Advisors', and 'Events'. The 'Events' section lists 'ICT Proposers' Day 2016' and 'Digital Assembly 2016'. The main content area displays the event 'Addressing the Big Data and Data Science Skills Gap' under the heading 'ICT 2015 Networking Sessions Programme'. The event details are: 'Room 5A, 22/10/2015 (09:50-10:35)'.

# Initial training is not enough

- Opening (or converting) masters in Data Sciences will not fill the Data scientists skills gap quickly enough
- **Solution: continuing education (lifelong learning) and learning on the job**
  - Example : At CNAM (70 000 “students” in all fields ) we have developed a professional certificate in Big Data analytics (27 ECTS), for graduated statisticians and IT engineers, which may be deployed in our network of 150 sites in France

Entreposage et fouille de données	STA211 9 crédits
-----------------------------------	---------------------

Bases de données documentaires et distribuées	NFE204 6 crédits
---	---------------------

Ingénierie de la fouille et de la visualisation de données massives	RCP216 6 crédits
---	---------------------

Projet	UASB03 6 crédits
--------	---------------------

# Technologies for massive training

- On line courses, MOOC
  - An issue: validation by credits, diplomas, accreditation?
- Need for cooperation and mutualisation of efforts
  - Cf. “European Data Science Academy”, a Horizon 2020 project





# Which role for learned societies?

