

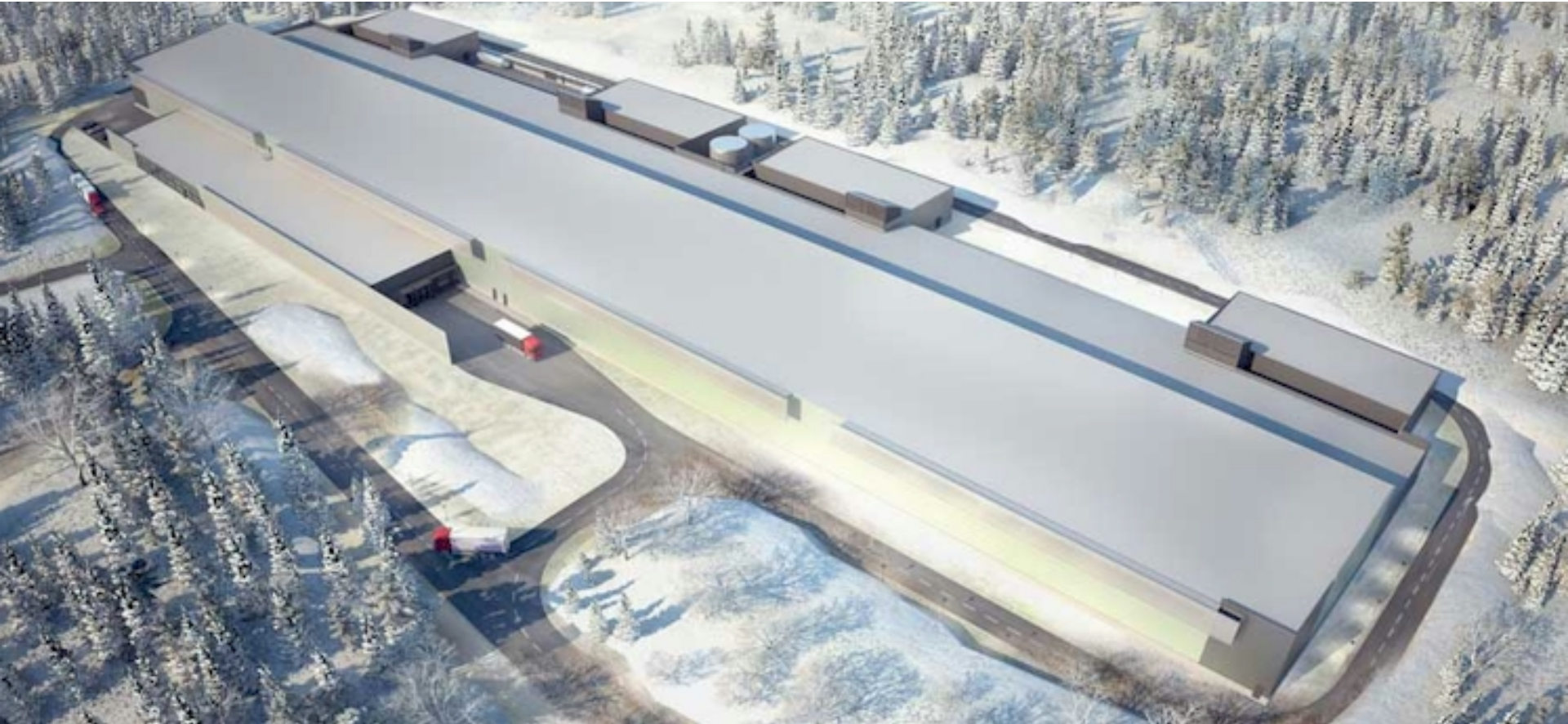
# **What is Big Data and Why Should We Care?**

Gianluca Demartini  
Lecturer in Data Science  
University of Sheffield  
[gianlucademartini.net](http://gianlucademartini.net)

# Data is huge

- Banks, city councils, governments, shops, etc.
- Facebook processes 750TB/day of data
  - 48k iPhones every day
  - 7PB of photo storage / month
- This requires computers (a lot of them!)

# Facebook Data Center (Sweden)







# Machines

- Google has around 900,000 servers (260 million watts) 0.01% global energy
- CERN Large Hadron Collider 180MW
- The flux-capacitor needs 1.21 GW (just 5xGoogle)

# Big Data

- Defined as **Vs**
  - **Volume**: Just about *size*, Giga, Tera, Petabytes
  - **Variety**: *Formats*, text, databases, pictures, excel
  - **Velocity**: *Speed*, 10 000 tweets per second, 2 000 pictures on Instagram per second

# Big Data Research in Sheffield

- Looking at data integration across sources

APRIL 9, 2012, 1:15 PM **MERGERS & ACQUISITIONS**

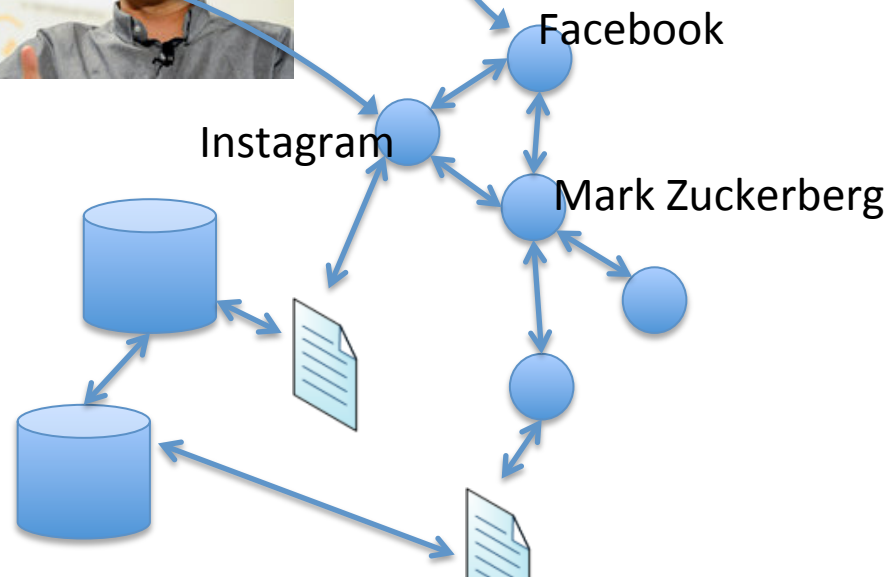
## Facebook Buys Instagram for \$1 Billion

BY EVELYN M. RUSLI

2:02 p.m. | Updated

Facebook is not waiting for its initial public offering to make its first big purchase.

In its largest acquisition to date, the social network has purchased Instagram the popular photo-sharing application, for about \$1 billion in cash and stock, the company said Monday.



# Should we care?

- This data is about us!
- **Data:** GMail, Facebook, debit cards, shopping fidelity cards, transport, mobile phones, ...
- **Usage:** Mortgage application, health insurance, car insurance



# Algorithms rule the world

- Some data must not be processed by people!
  - GMail content is processed by computers to decide which advertisement you see on the Web



# Algorithms rule the world

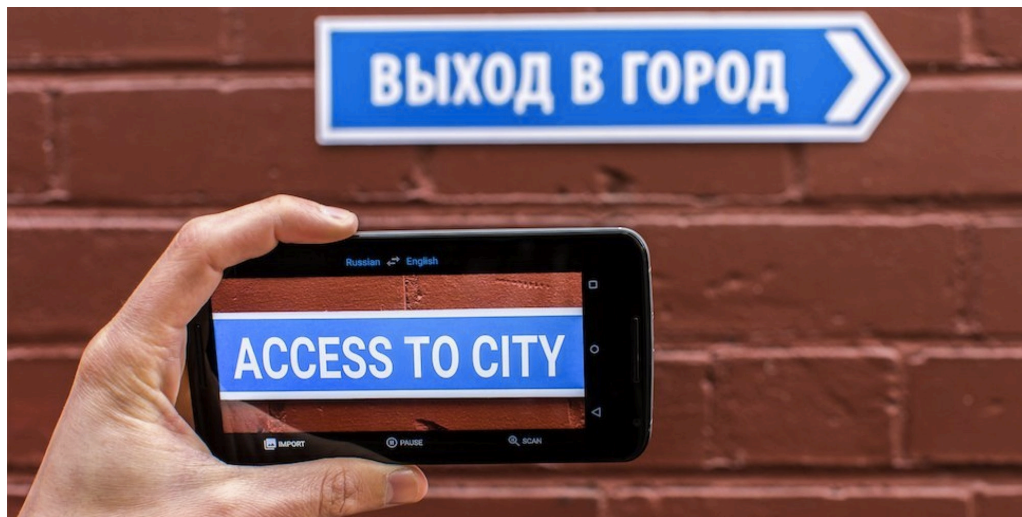
- **Uber** prices are decided by a software programs
  - The boss of Uber drivers is a computer
  - It decides how they work and how much money they make
- Computers know a lot about people but not the other way around



U B E R

# Is it all bad?

- Duolingo: Data-driven foreign language learning
  - What is the best way to learn a language depends on your native language
- Language translation



# Data Science

- “Data Scientist: The Sexiest Job of the 21st Century”, in Harvard Business Review
- Companies want **data-driven decisions**
- Graduates from the MSc Data Science in Sheffield go work in:
  - Telecommunication data analysis
  - Cancer research
  - Housing market

# Is your data ready?

- Decisions about you and your life will be driven by data!
- Are you in control of which data is being collected?
- Are you aware of how the data is stored and processed?
- Do you know how decisions are taken? Do you want to know?