TSSG

Challenges of Networking, Security and Big Data in the IoT

Bernard Butler

Emerging Networks Laboratory, TSSG, WIT

Waterford Institute *of* Technology
INSTITIÚID TEICNEOLAÍOCHTA PHORT LÁIRGE

Overview

1. TSSG: Who we are
   - Background
2. Infrastructure Research
   - Networking
   - ICT Convergence
   - Security and Privacy
3. Big Data
   - Some definitions
   - Exercise
   - The Context and History
   - Expertise and Roles
   - Is the IoT and Big Data always a force for good?
4. Conclusions

Overview

- Founded 1996 by Prof Willie Donnelly and Eamonn de Leastar
- Funded research, industry partnerships, spin-out/in
- Research staff (scientists, engineers, designers), Interns and postgrad students
- Broadened to ICT centre of excellence in SE region
- International links: EU H2020, SFI Research Centres, Tech Committees
- 500 partners (140+ Irish)

Research highlights

- 1996-2009: telecoms network management, billing, making content available, personalisation, context awareness
- Consumer-facing research led to *mobile apps*, via App Store, Google Play, etc.
- Infrastructure research led to cloud and edge (IoT) computing
- New research themes include
  - zero-touch network management (adaptive, autonomous)
  - high data rate, low latency communication
  - smart agriculture: precision dairy, animal health
  - Security and privacy of pervasive computing, social media
  - the Brain Initiative: modelling neural processes
  - the Connected Human: in-body networks

Network Evolution

Growing Consumer Requirements

- More reliable connection
- Better Quality of Experience
- Lower usage charges
- Emerging needs

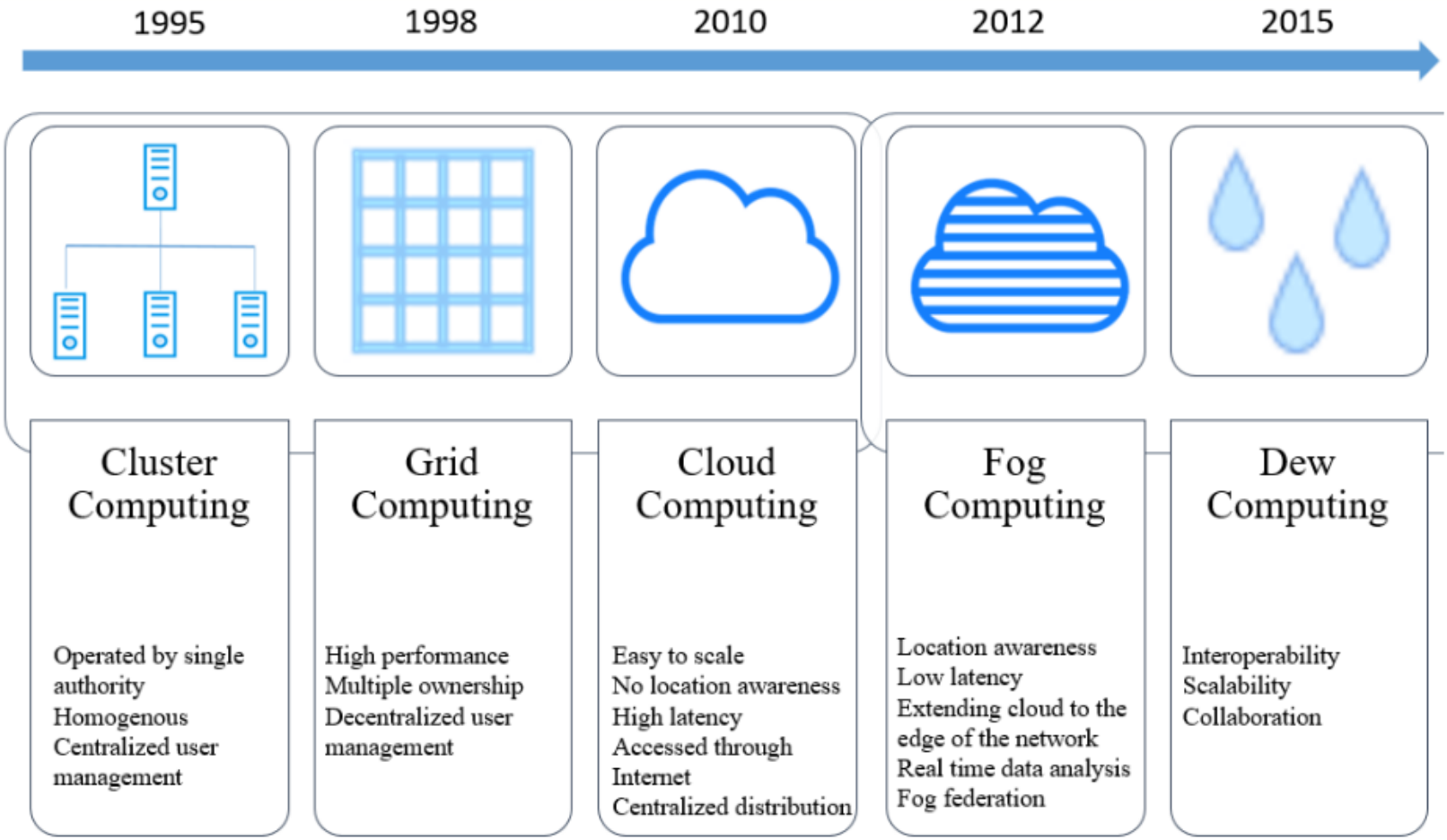
Growing Operator Requirements

- Increase flexibility to meet demand
- Make more efficient use of existing resources
- Reduce CAPEX and OPEX
- Provide for new requirements

Growing Domain Influences

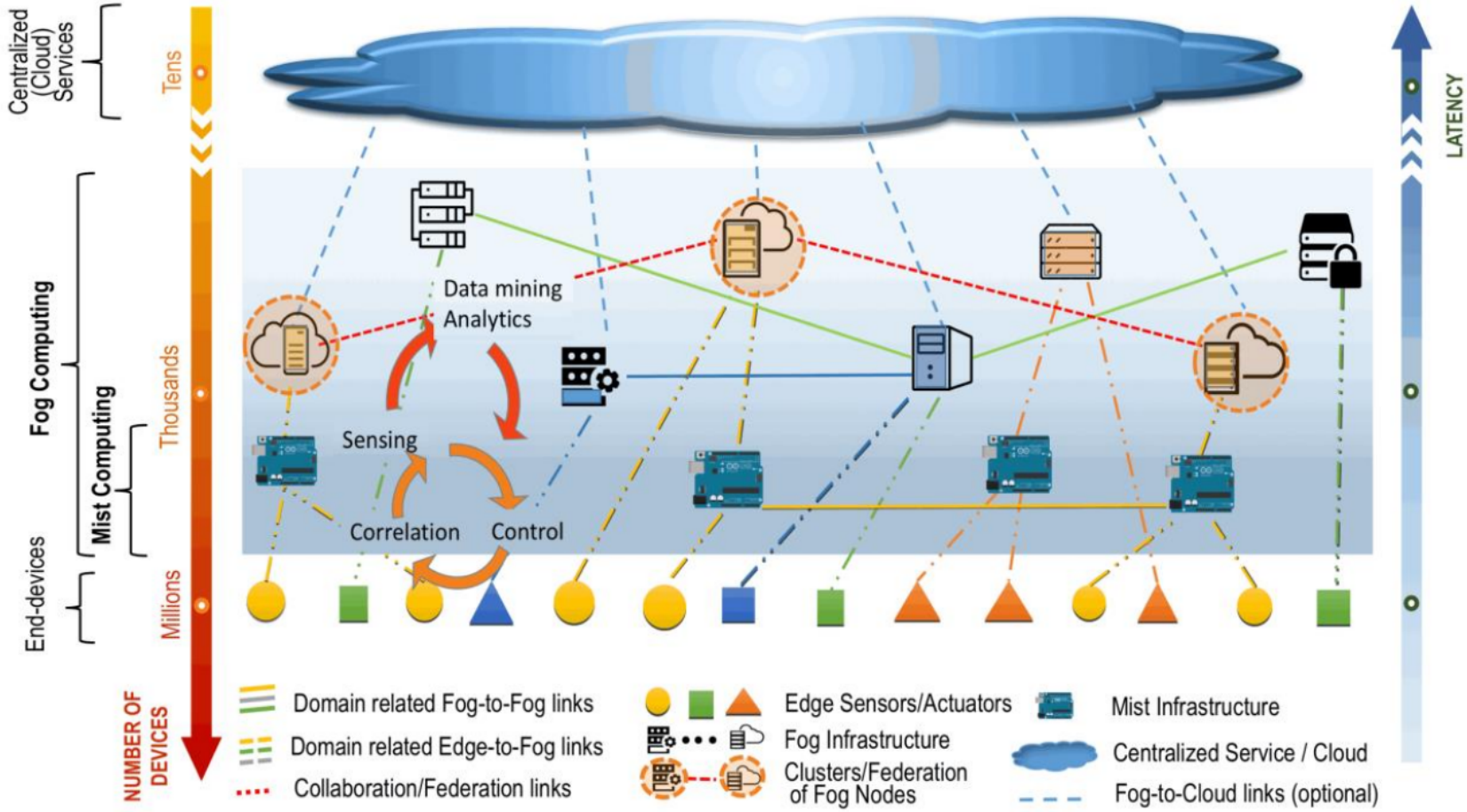
Transport, Health, Agriculture/food, Media, ...

Evolution of computing paradigms



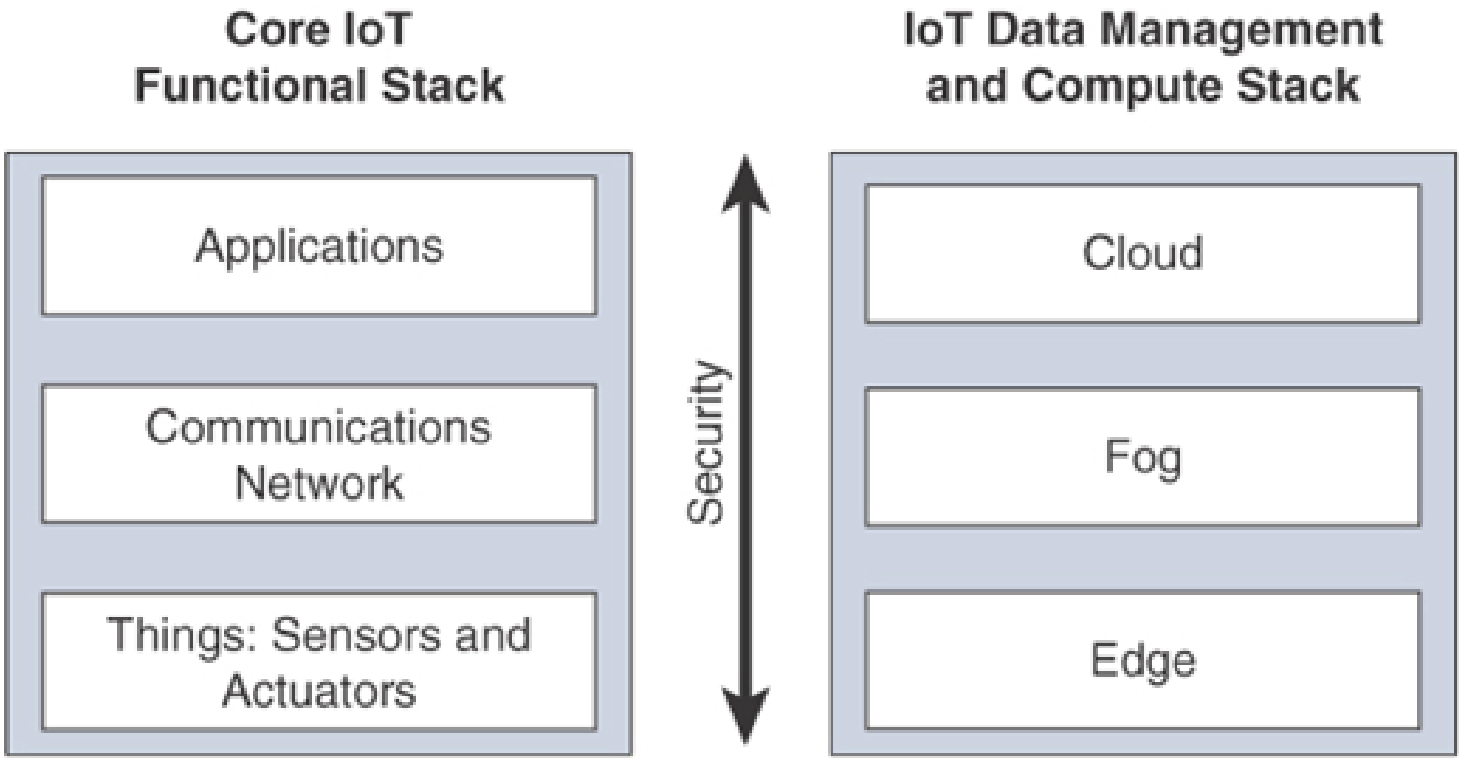
Source: Loncar, Petra (2018). Data-Intensive Computing Paradigms for Big Data

Cloud, fog and Edge



Source: NIST SP 500-325. Fog Computing Conceptual Model

Internet of Things as Two Stacks



Source: Yogesh Malik, medium.com. Sep 2017

Security and Privacy

- In pairs, identify 3 ICT security "stories" of the past 5 years.

Selected definitions

Info security: maintaining *appropriate access* by protecting the confidentiality, integrity and availability of data and its supporting infrastructure (Lundgren 2017).

IoT: Network of pervasive connected devices that exchanges data from embedded sensors, with supporting infrastructure and services. (Various)

Big Data: High volume, velocity and/or variety information assets that require new processing to enable enhanced decision making, insight discovery or process optimization. (Gartner 2012)

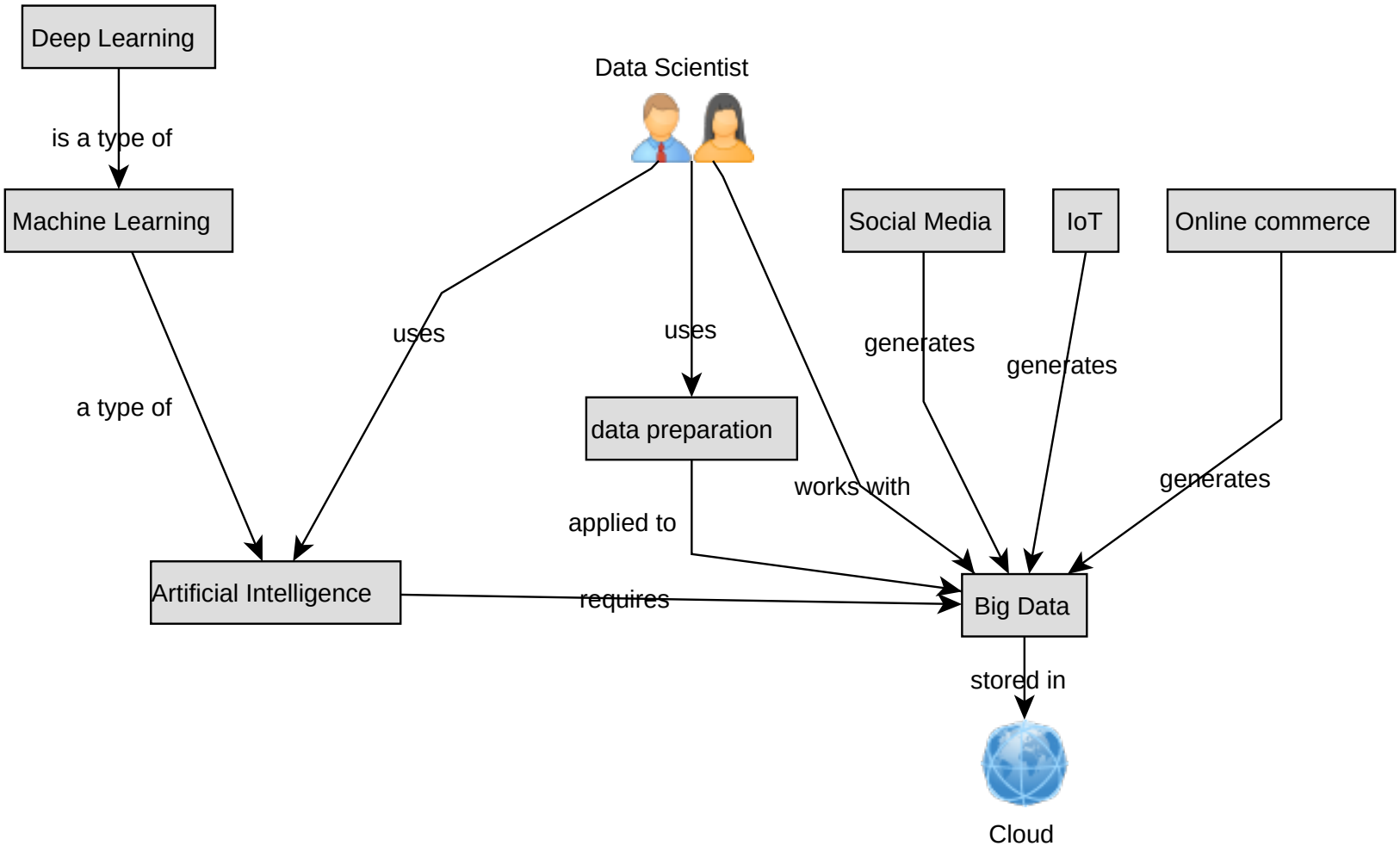
Data Scientist: ask the right questions, {generate} and consume the results of analysis of Big Data effectively. (McKinsey 2011)

Artificial Intelligence: capability of a machine to imitate intelligent human behavior (Webster 2017)

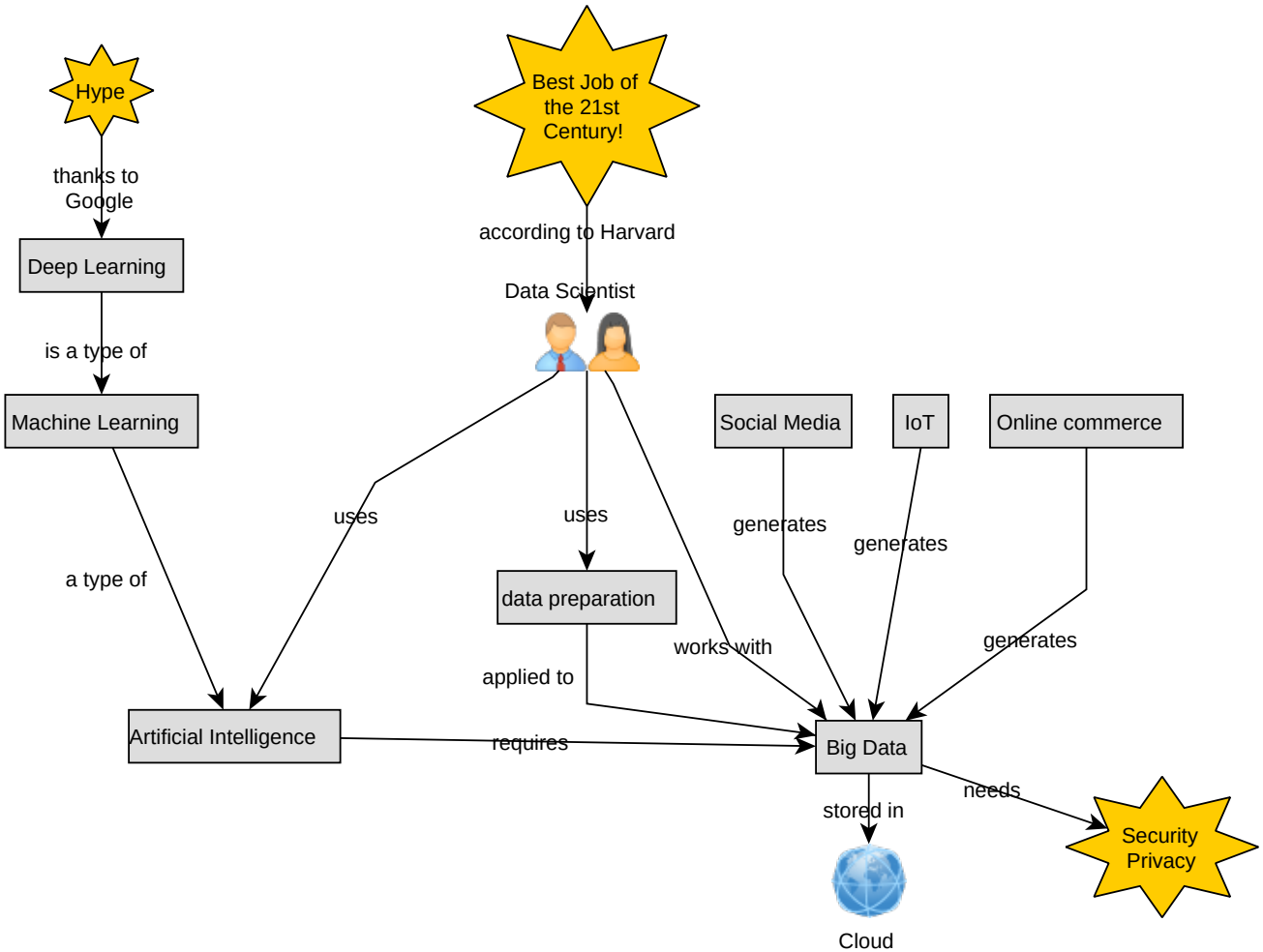
Machine Learning: Branch of computer science {and related fields} that gives computers the ability to learn without being explicitly programmed. (Samuel 1959)

Deep Learning: Use of very large neural networks with many layers of "neurons" that can be trained to generate robust models of their input, whose classification performance scales with the amount of data supplied. (Various)

Relationships between terms

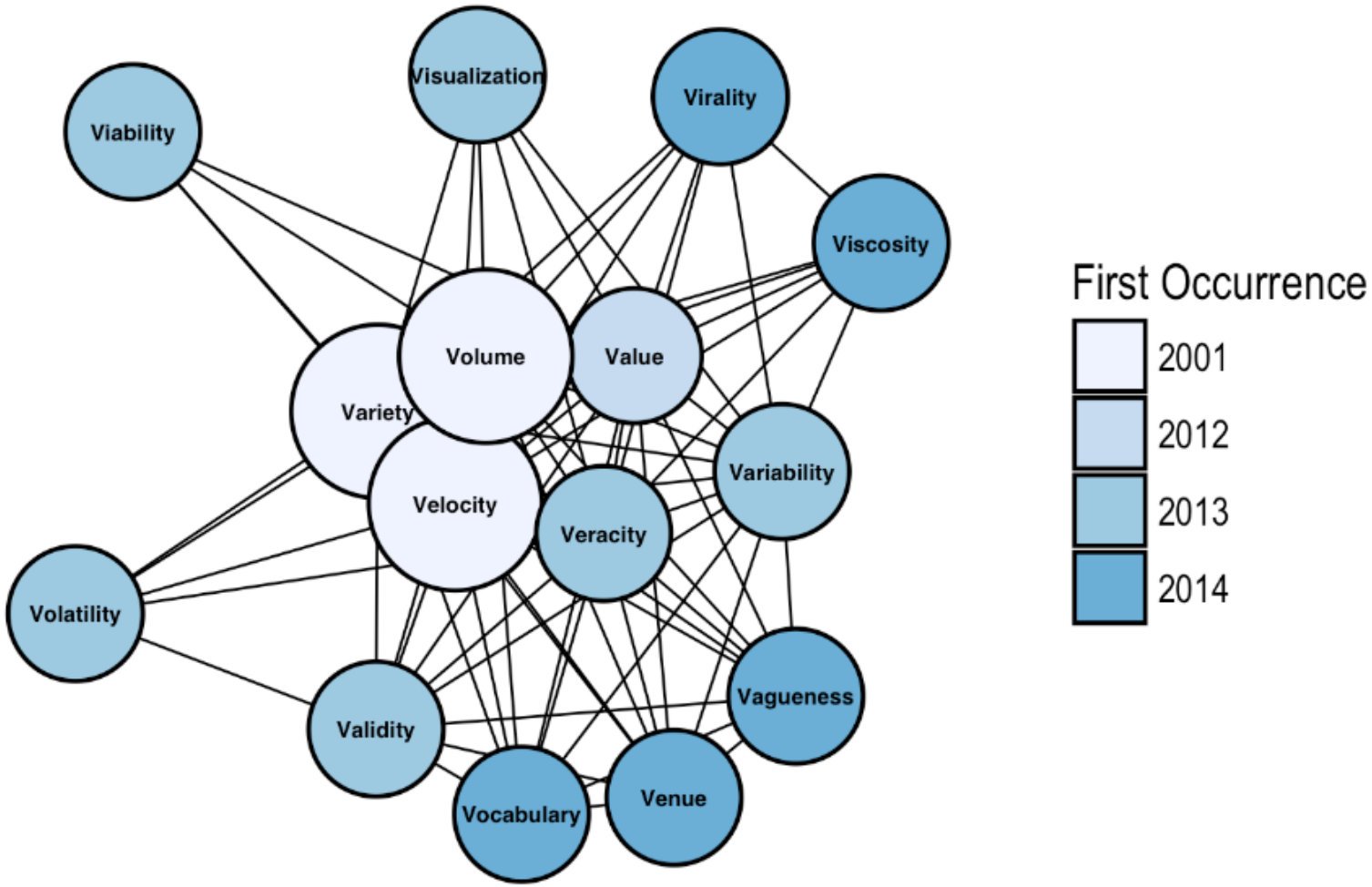


Annotated Relationships between terms!



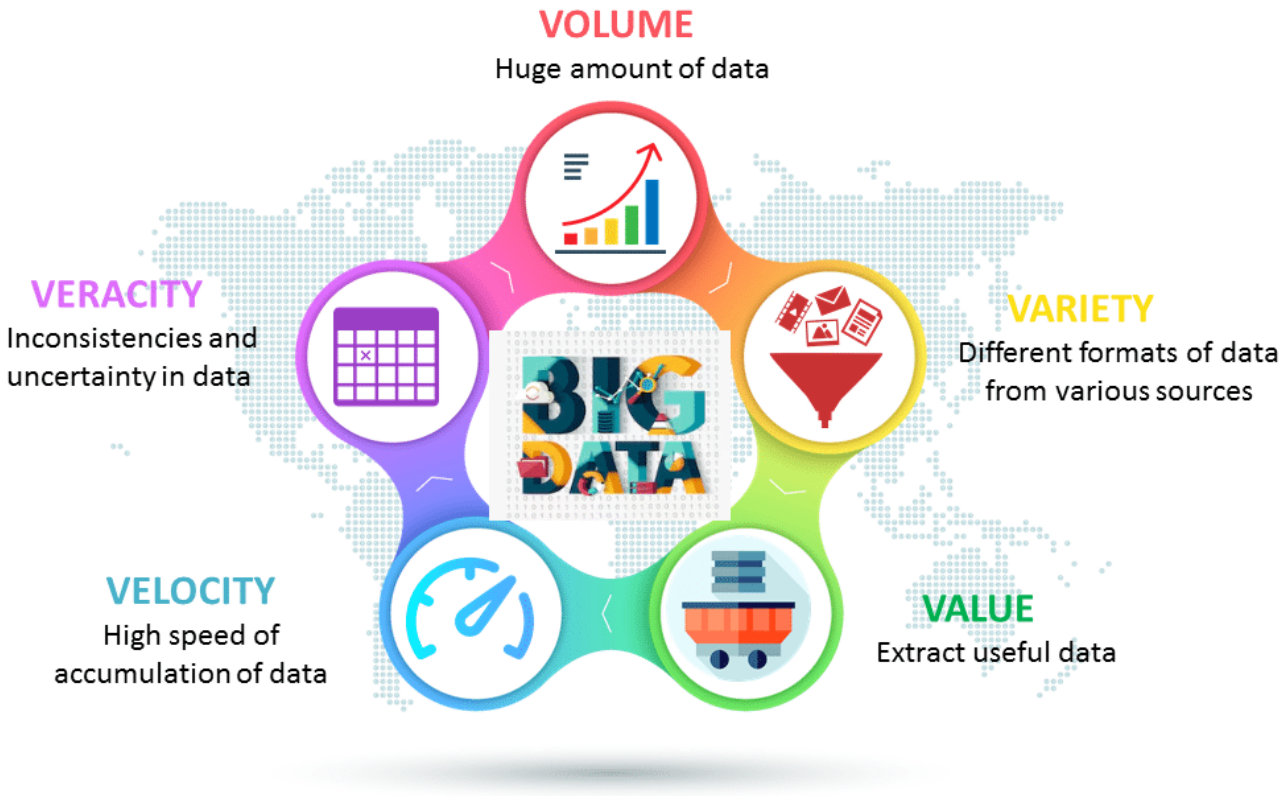
The V's of Big Data

First there were 3: Volume, Variety, Velocity.
Now...



Source: Elder Research 15V's - Tom Shafer

Most popular variant: 5 V's

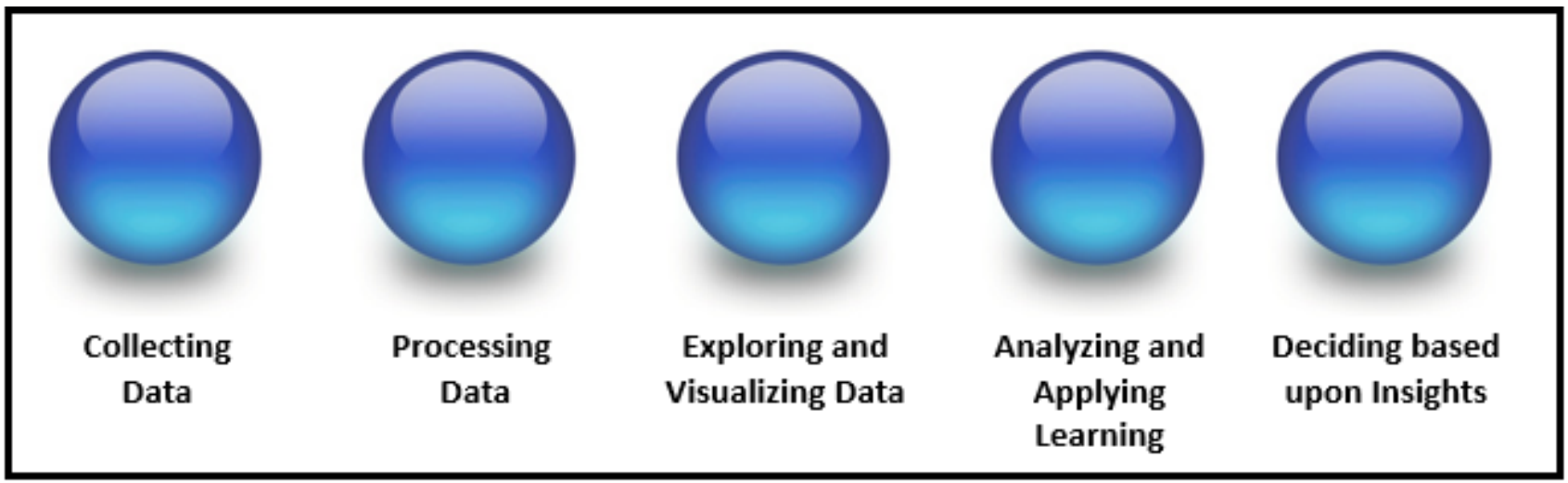


Source: https://www.edureka.co/blog/what-is-big-data/

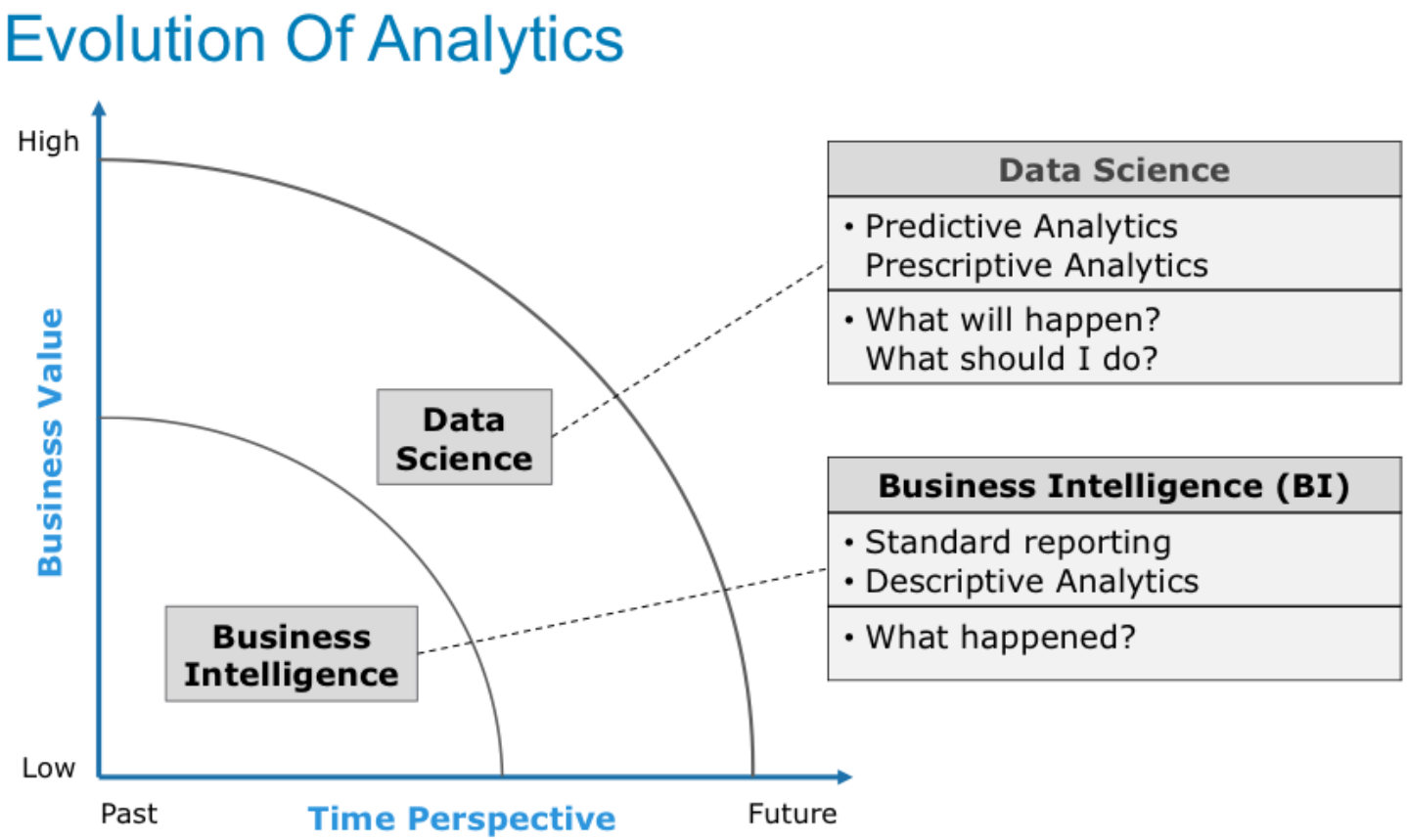
Interlude: Examples of Big Data

Exercise

In pairs, please consider (real world) processes generating *Big Data*. Can you come up with 3 examples in 2 minutes?



The growing scope of Data Analytics



Source: Bill Schmarzo "Big Data MBA" Course Curriculum

Prehistory, or more than 10 years ago...

Data Generation

- Transactions (bank, retail)
- Activity, e.g., texts
- Basic e-commerce

Data Processing

- Databases, SQL, stored procedures
- Consultants, system integrators
- Proprietary statistical software

Data Analysis

- Reporting: looking back
- Descriptive statistics
- Simple plots

The first (batch) wave: 2007-2011

Data Generation

- As before...
- Web activity: comments, etc.
- 360degree view

Data Processing

- As before...
- NoSQL
- hadoop ecosystem (batch analytics)

Data Analysis

- As before...
- Personalisation and recommendation
- Predictive Analytics

The second (streaming) wave: 2012-2015

Data Generation

- As before...
- Social Media!
- IoT (early adopters)

Data Processing

- As before...
- Apache Spark
- R vs. python

Data Analysis

- As before...
- Data understanding
- Weak AI: assistants, etc.

The current (machine) wave: 2016-?

Data Generation

- As before...
- Machine-generated (e.g., fake news)
- IoT (mainstream)

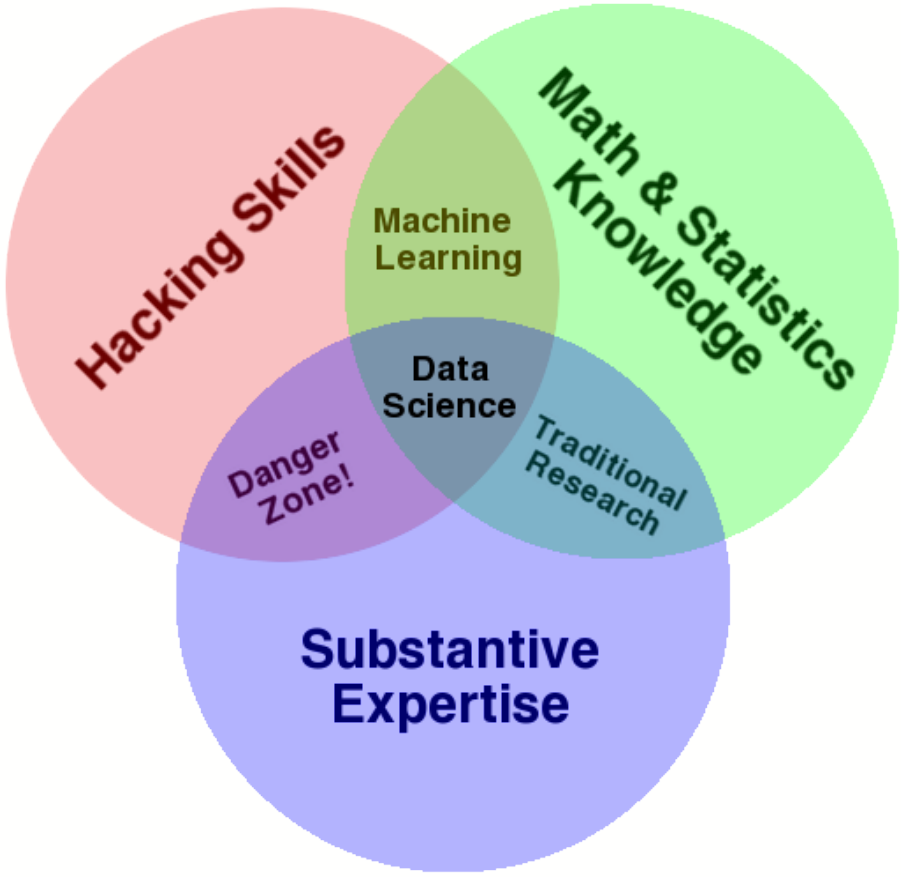
Data Processing

- As before...
- Microservices: move function to data
- Decoupled databases with schema-on-read

Data Analysis

- As before...
- Deep learning inflection point
- Visualisation

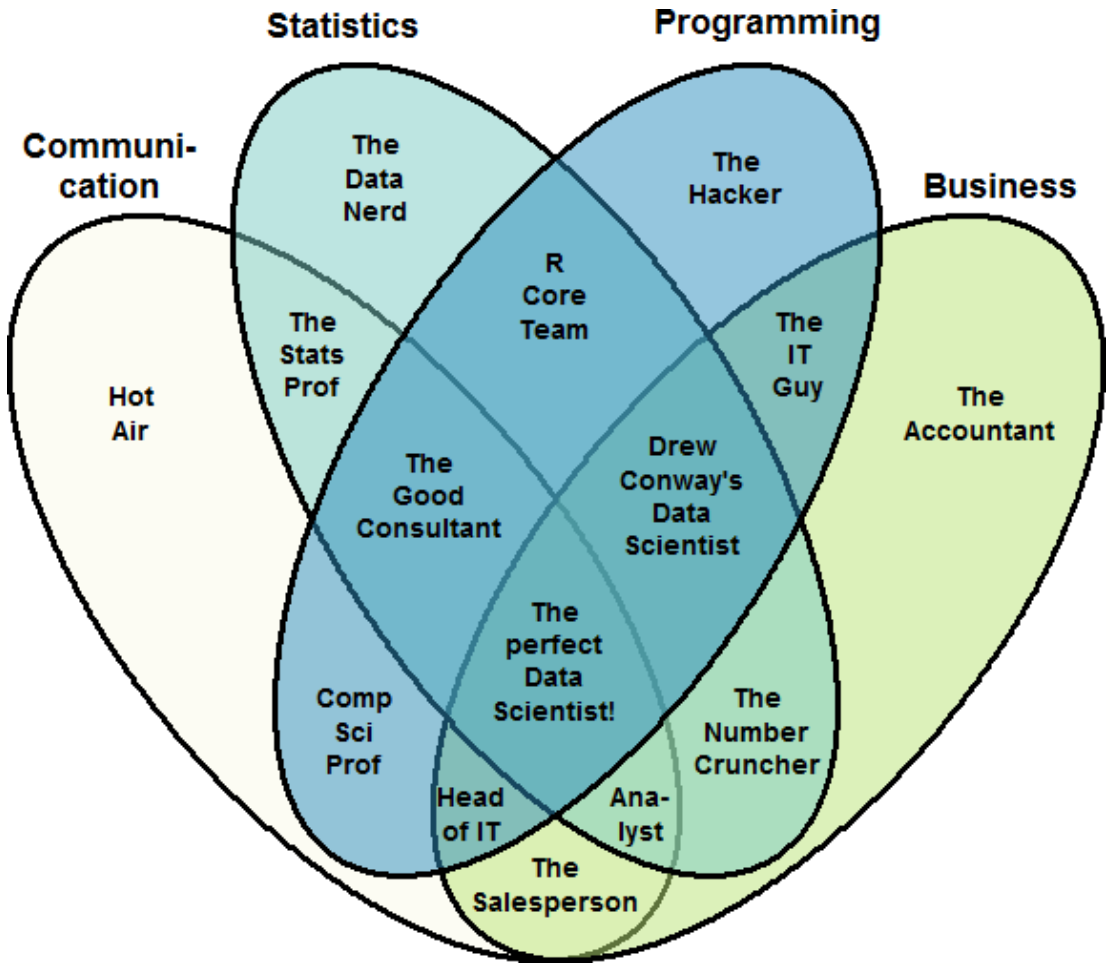
Drew Conway's 3-set Venn Diagram of Data Science Expertise



Source:

`http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram`

Stephan Kolassa's 4-set Venn Diagram of Data Science Expertise



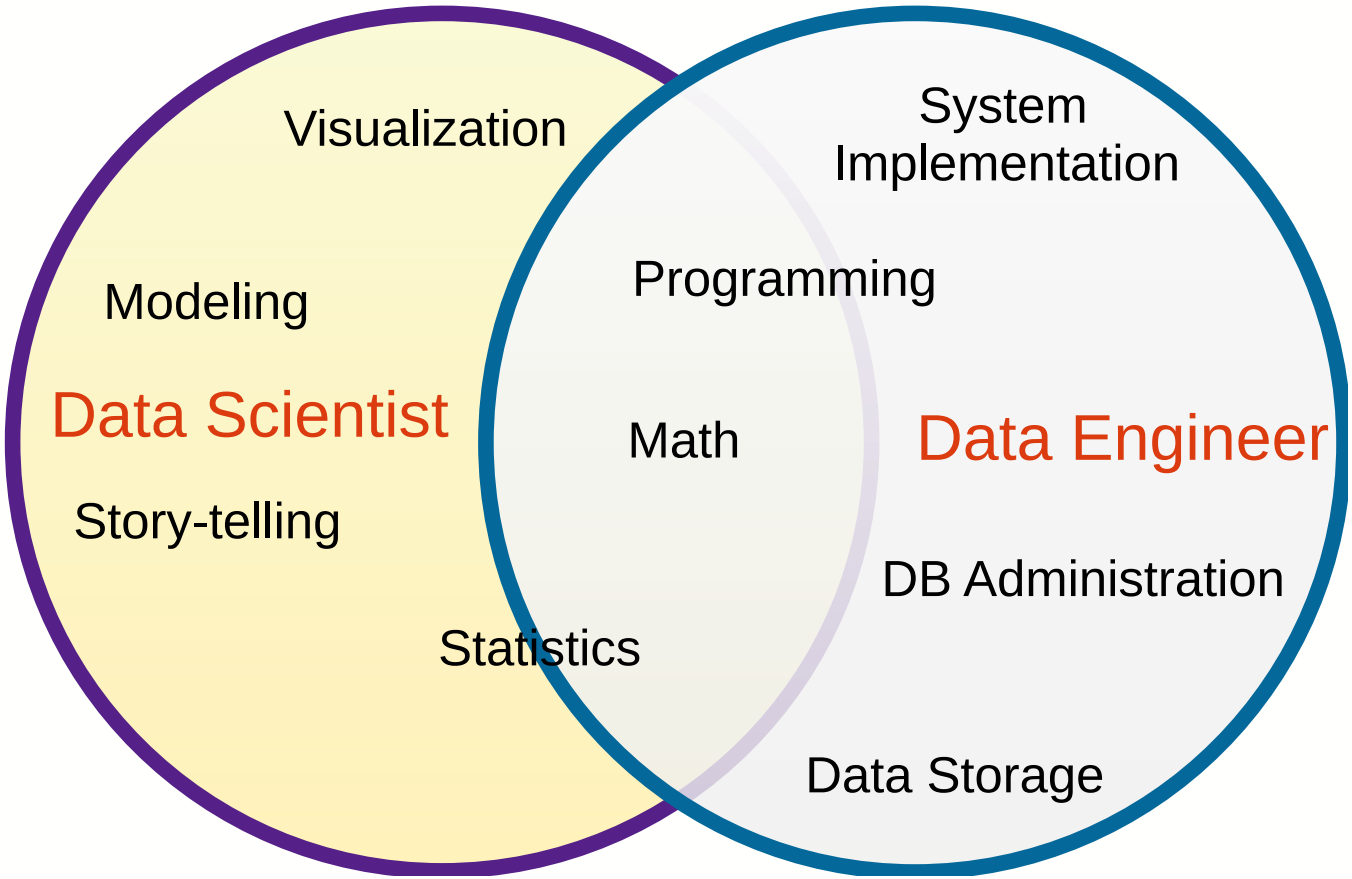
Source: `https://datascience.stackexchange.com/a/2406`

Gartner suggests the need for a *Citizen Data Scientist*

Source:

`http://www.kdnuggets.com/2016/03/cartoon-citizen-data-scientist.html`

Data Scientist vs Data Engineer



Source:

`http://101.datascience.community/2014/07/08/data-scientist-vs-data-engineer/`

Also the traditional roles of *Data Analyst* and *Software Engineer*. . .

Complete the following disadvantages of IoT and Big Data

m___ s__v__l_____

i___t__y _h_f_

d___c_ b__n___

d____l _f __r__c_

b__s

l___ o_ t___s__r___y

And those disadvantages are. . .

mass surveillance

identity theft

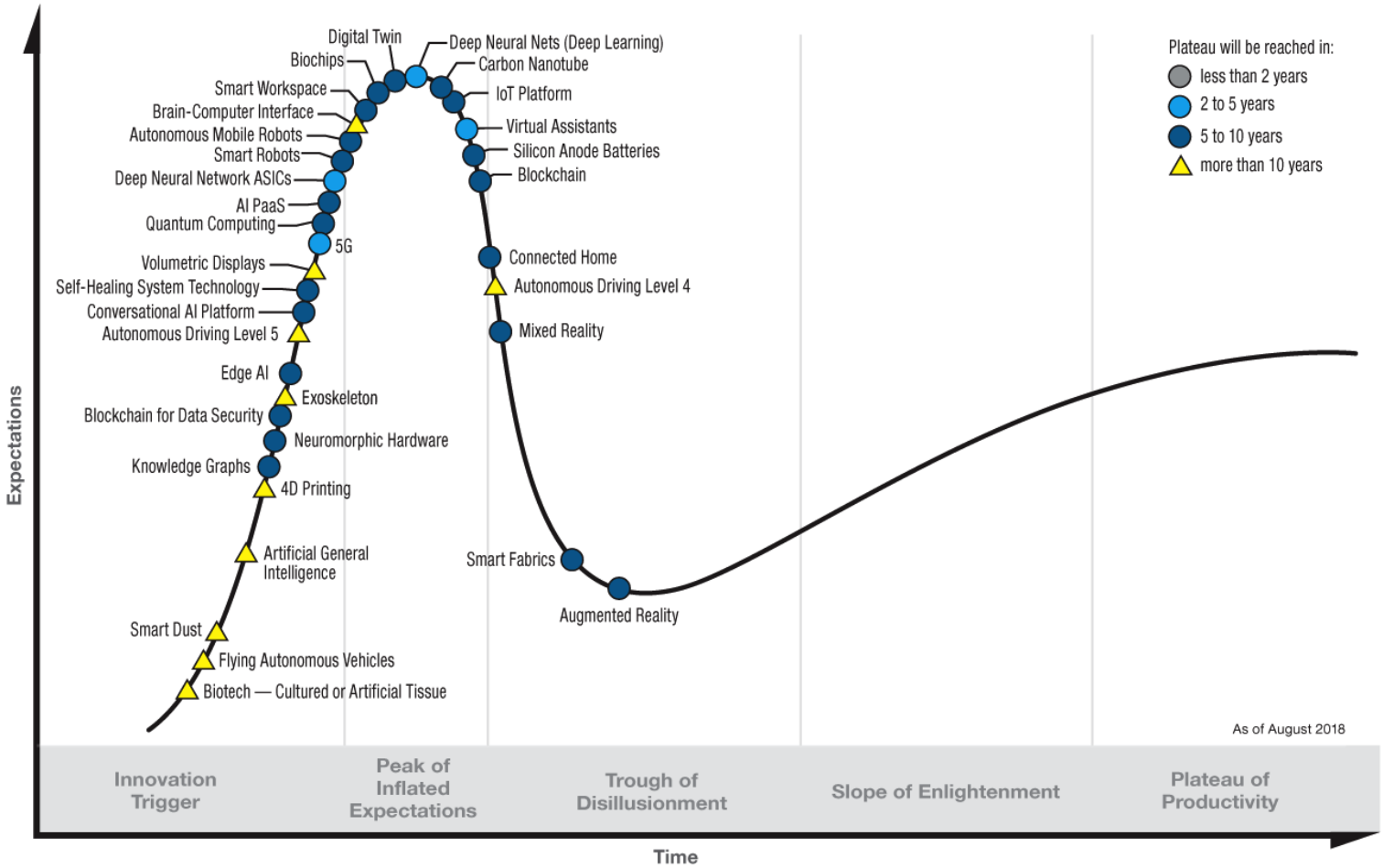
device botnets

denial of service

bias

lack of transparency

Where will your career take you?



Source: Gartner Hype Cycle for Emerging Technology, August 2018

Conclusions

- Computing is becoming more interdisciplinary
- Research challenges: how to do things better: faster, more accurate, less energy, ...
- Societal challenges: how to use these new devices, services, interactions, ...; ethical use
- Many computing jobs to be filled - so good luck!

Thank You

Contact:
Bernard Butler,
Research Fellow, TSSG
Waterford Institute of Technology, Ireland
t: +353 (0)51 84 5695
e: bbutler@tssg.org

Waterford Institute *of* Technology
INSTITIÚID TEICNEOLAÍOCHTA PHORT LÁIRGE