

An overview of Big Data

Dr Philip Woodall

Senior Research Associate
Institute for Manufacturing
University of Cambridge

Contents

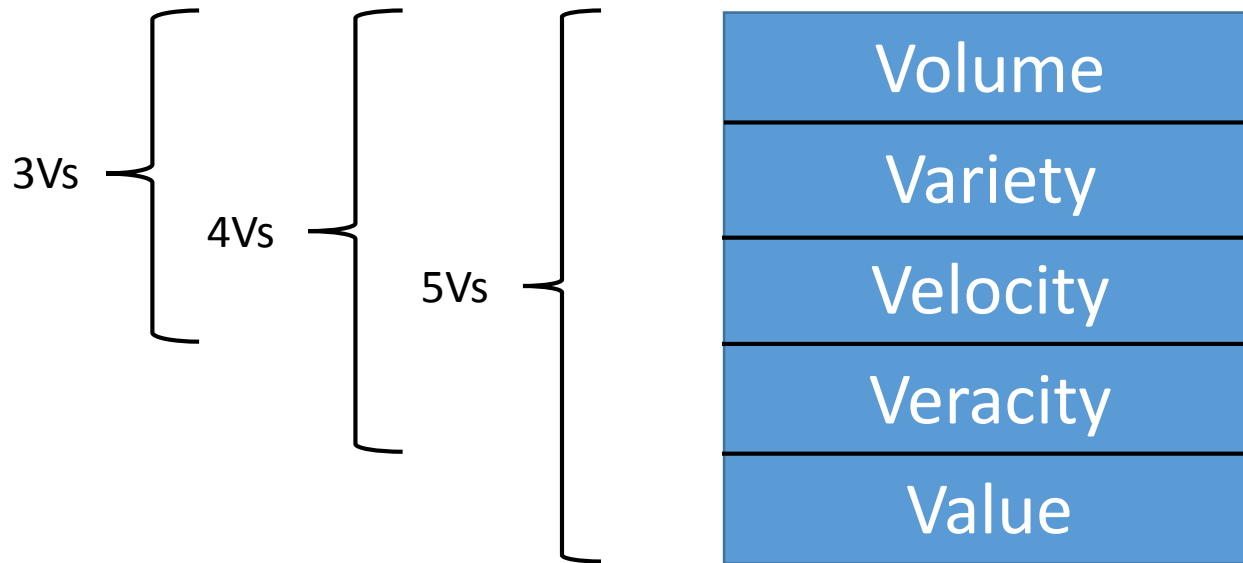
Big Data – What is it?

Big Data – Opportunities

Prediction Capability – Inspirational examples

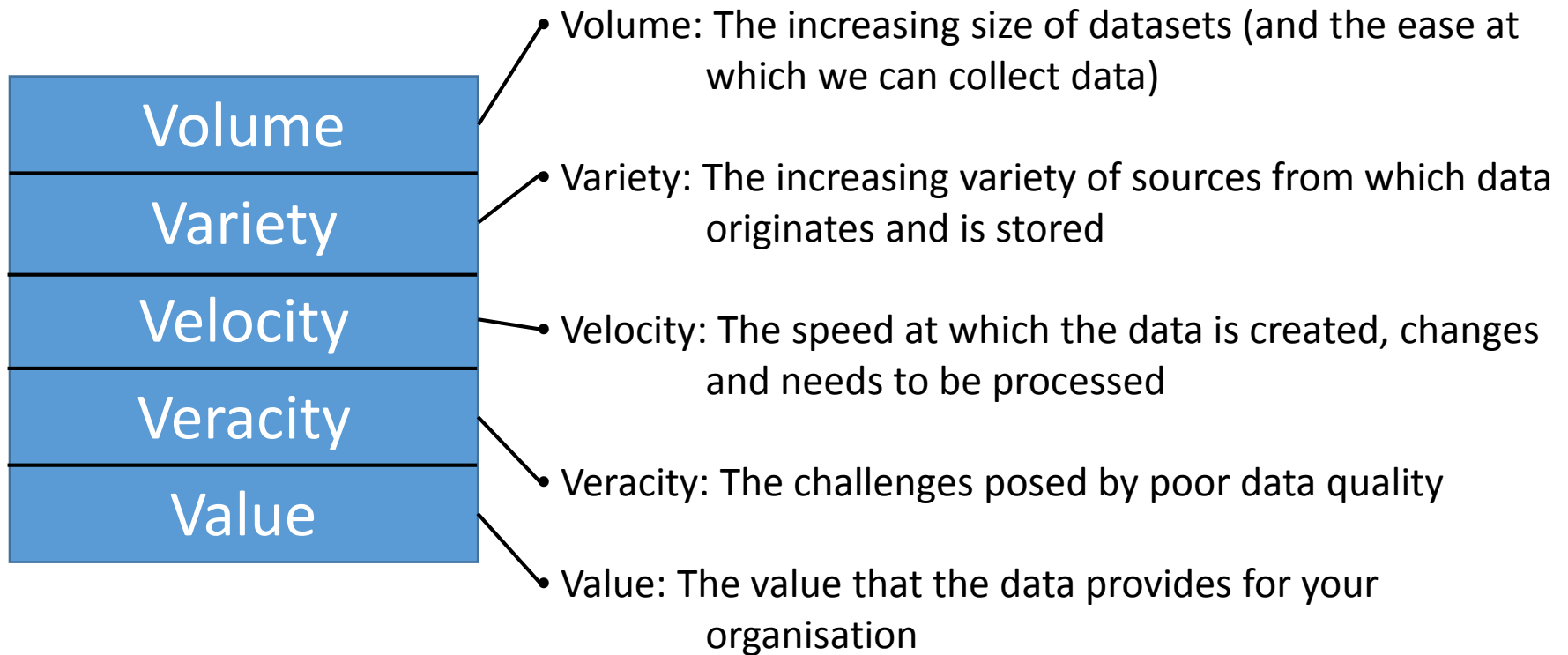
Big Data – what is it?

- The “V”s definition:



Big Data – what is it?

- The “V”s definition:



Volume

- Data creation:
 - 1 flight from a Boeing 787 can generate up to ½ terabyte of data
 - More data from customers - they converse on a global scale
 - E.g needs/desires; product reviews; when, how they use the product
 - Enabled by automated capture devices: sensors, RFID, Bar Codes, point of sale capture devices, GPS Location data, smart phones...
 - Online platforms (iTunes, Netflix, Amazon, Google) all recording what, when, where, how long, how much, who...
- Speed and scale of data sharing:
 - Product reviews accessible instantly and globally
 - posted, tweeted, “liked”, recommended
 - enabled by smart phones, social media, online platforms

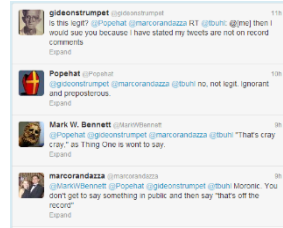
Variety



Sensors



Public data

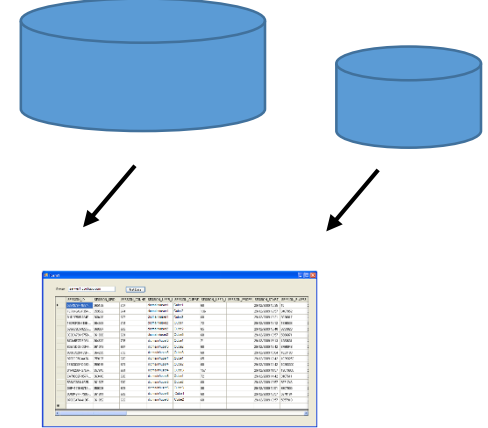


Customer Type	Customer ID	Registration	Location	Registration Date
...



...
10000000	1740-01-01	00	00	00	00	00	00	00

Internal data



Different data formats and media

Different data syntax and semantics

Does the data refer to the same customer, company, supplier?

Velocity

Time

Data collection

Data pre-processing
(cleansing, integration etc.)

Data processing

Data
interpretation

Decision making

Action

Velocity

Time



Data collection

Data pre-processing
(cleansing, integration etc.)

Data processing

Data
interpretation

Decision making

Action

Deadline

Can all these steps keep pace
with the volume of data that
needs processing?

Can action be taken
before it's too late?

Veracity

Time

Data collection

Data pre-processing
(cleansing, integration etc.)

Data processing

Data
interpretation

Decision making

Action

Poor data quality costs the
US economy approx.
\$3.1 Trillion per year.

Garbage in garbage out

Don't underestimate the
effort to pre-process data:
80% of the project time is the
widely estimated value

Value

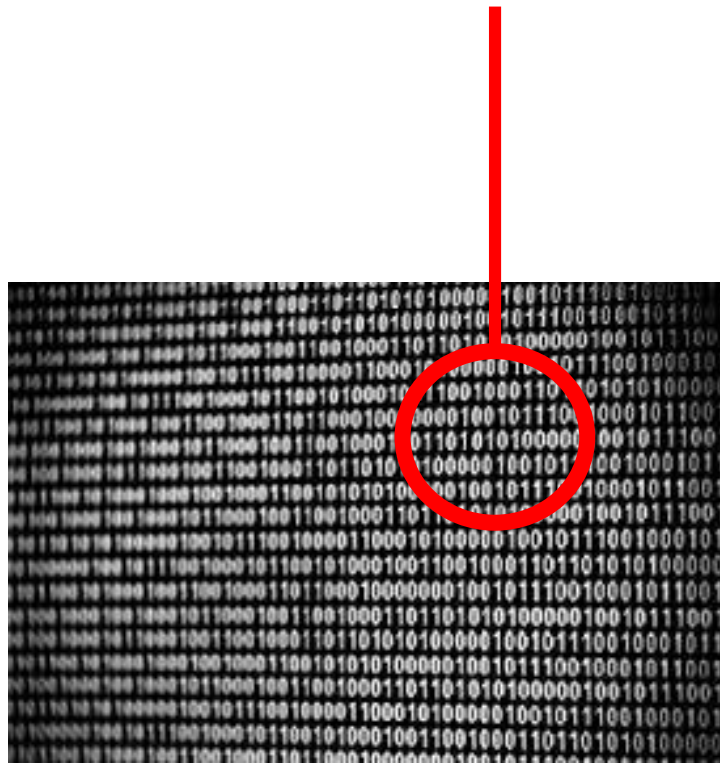
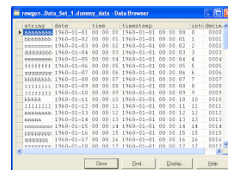
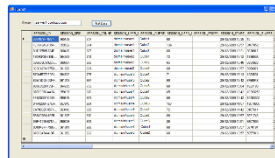
What data is most valuable to your organisation?

Data is not actually free:

Time to collect, pre-process, etc.

+ data management costs

+ Risk if released/hacked



Big Data – Opportunities

- “Big Data” also refers to the new opportunities to use data to inform decision making and discover insights for the business

- Some key terms:

Data Analytics, Business Intelligence, Predictive analytics, Data Scientist

- Essential skills:

—statistics (including data mining, machine learning)

—data manipulation (including programming/scripting)

—data storage (databases, SQL, Hadoop, data warehouses)

—data awareness (what is available and what can be made available)

—business domain knowledge

Prediction capability – Inspirational examples

- Big data to cut delays on construction sites by predicting what's under the ground before digging
- Data has been integrated from:
 - Defra, English Heritage, Scottish Natural Heritage, Ordnance Survey, the Land Registry and many more.
- This could save companies millions by avoiding ancient burial grounds etc.
- IET E&T magazine, March 2015

Prediction capability – Inspirational examples



TARGET

Questions

Dr Philip Woodall

Senior Research Associate
Institute for Manufacturing
University of Cambridge