## 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

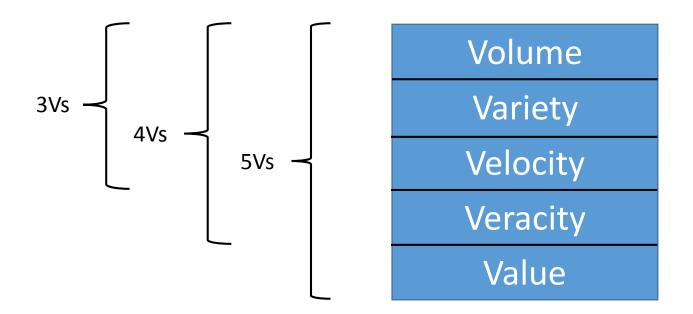
Prediction Capability – Inspirational examples





## Big Data — what is it?

The "V"s definition:

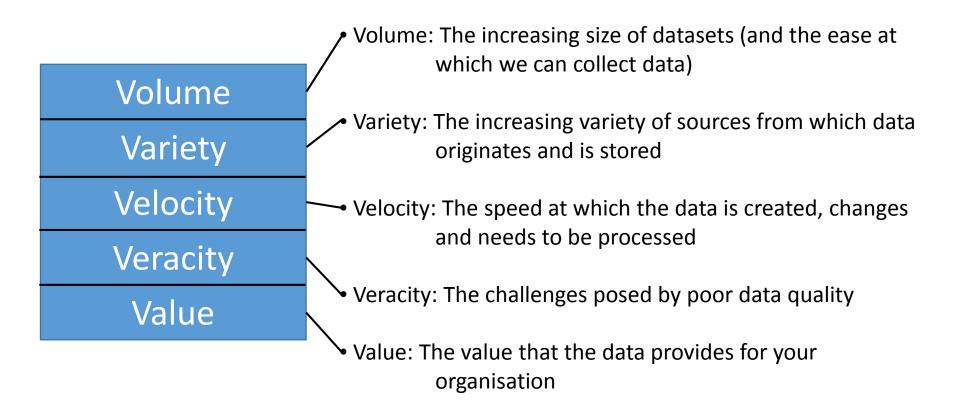






## Big Data – what is it?

The "V"s definition:







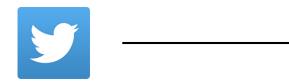
#### Volume

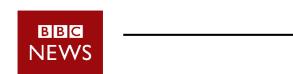
- Data creation:
  - 1 flight from a Boeing 787 can generate up to ½ terabyte of data
  - More data from customers they converse on a global scale
  - E.g needs/desires; product reviews; when, how they use the product
  - Enabled by automated capture devices: sensors, RFID, Bar Codes, point of sale capture 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



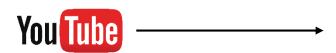


### Variety











Sensors



#### Public data



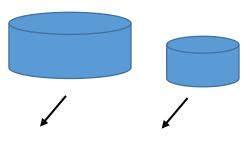


Request Type Cal Cass Records	Bequest la	Grade 7	Requesting Officer Materialist Police	Request Date CO-Jan-37			
Feature No.	Start Bate 14 Noville	Start Time (0:00:0)	Tred Date 14 har-05	23:59:50			
Part Des	-	Memory Type	A Digital	B Digits	Direton	G	Ste Name
14/00 CG	(M17.54	(Motors Concenting	077 ex xxx347	25	00:01:50	0547	Etringham Street
1879=05	1594.64	Decisio Levelingting	- 10	077ex xxx347	00.00.50	5943	Gruber Road Water Tower
14 No.05	1107 61	Mobile Cooperating	077 ec 200347	11	00:00:07	1643	Gruber Road Water Tower
		Mobile Originating	027 ex xxx347	99	00:00:24	5567	Gruber Road Water Tower
14-Nov-06	1109.10	Michie Connattre	577 m xxx34?	316	00 00 23	1061	Gruber Road Water Tower
1876m/5		Motor Colpanies	\$27 to sex347	33	00:00:54	1643	Orober Road Water Tower
Street.	125.3	Motors Organitry Motors Organitry	077 ex 221347	333	00.02.48	85a7	Eltrinoham Street
147ke-05		Water Terminating	377 ex xxx347	. 11	00 00 36	0547	Etringham Street
14-74to-CK	1521 (			077xxxx047	00:00:36	85a7	Etropham Street
14.Nex-06				077xx xxx347	00:01:12	1609	GNG House
18-No. (K)	1766 (4	Mates Cognitive	107 er 201347	C77xx xxx347	00 00 29	05x7	Eltropham Street
			101.61 137.761		00:00:09	65-47	Etrophan Street





#### Internal data



Learning Co.	STREET, STREET	2117 CL 6	STREET, STREET	22117 CR12	OFFICE LITTLE PROPERTY.	III DESIGNATION AND ADDRESS OF THE PERSON ADDRESS OF THE PERSON AND ADDRESS OF THE PERSON AND ADDRESS OF THE PERSON ADDRESS OF THE PERSON ADDRESS OF THE PERSON AND ADDRESS OF THE PERSON ADDRESS OF THE PERSON ADDRESS OF THE PER	200.70.0.00
20/8/2010	2013	O.		Cabri	4	25/10/10	
number to	5000	24	description.	Carbon St.	19	240,000,000	A701
SALES SALES	980	45	Service of	300	16	THE OWNER OF	000000
1000000WW.	717	100	COLUMN TO SERVICE STATE OF THE	2VM	et .	7915 THE CAT	scwee-
			STREET, ST.				
			foreither!				

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



Data collection

Data pre-processing (cleansing, integration etc.)

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

I I

Deadline

Time

Can action be taken before it's too late?

interpretation

Decision making

Data processing

Data

Action





#### Veracity

Data collection

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

Data processing
(cleansing, integration etc.)

Garbage in garbage out

Data interpretation

**Decision making** 

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

Action





#### Value

Data is not actually free:

Time to collect, preprocess, etc.

+ data management costs

Risk if released/hacked





APPROPRIES	SPRING SPE	HUNGLIS C	ORDER LAND	APPENDING THE PERSON	OFFICE LATER	STATE OF THE PERSON.	SHIP CHAP	March 1 1989	
				Calcula					
		63							
1093/86-		2.8	2075 19240	0.00	79		29 O. VIII 1: O		
		45	SHEET HARMY		10		21GVBUR		
			and the						
		G/			es.				
DESCRIPTION.							26 OLYMP II D		
					100				
		56			74		2910/2009 1142		





What data is most valuable to your organisation?







## 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







## Questions

Dr Philip Woodall

Senior Research Associate
Institute for Manufacturing
University of Cambridge



