# Users' Guide to Big Data The Poets' Guide

September 2016







- 1. Oxford English Dictionary: data of a very large size, typically to the extent that its manipulation and management present significant logistical challenges
- 2. McKinsey (2011 study): datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze
- **3. Gartner:** high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation
- 4. SAS: a term that describes the large volume of data both structured and unstructured that inundates a business on a day-to-day basis.... big data can be analyzed for insights that lead to better decisions and strategic business moves
- 5. John Henry (Maiden): data, typically including structured and unstructured, of sufficient size to require advanced tools and non-standard modeling techniques
- 6. Susan Athey (Stanford): it's not just the data; it's not all new; the whole is greater than the sum of the parts; it's crucial, transformational, and existential





#### How much do you have, how fast can you use it, how many and what types do you have?

- 1. Volume: how many terabytes, petabytes, exabytes, zettabytes, yottabytes, brontobytes or gegobytes of records, transactions, tables, files, videos, etc.
  - A gegobyte is 1,000,000,000,000,000,000,000,000,000 bytes
- 2. Velocity: batch, near time, real time, streaming
- 3. Variety: structured, unstructured, both

#### The fourth V is Veracity: is it accurate?

• This is important for small data sets as well, but can be harder to confirm/validate for big data or quickly changing data

### Big Data is REALLY BIG





Source: The Future of Cognitive Computing, Andrew Trice, November 23, 2015

### Where is This Data Coming From?





#### From 2014-2020, IoT grows at an annual compound rate of 23.1% CAGR



50

MAIDEN HOLDINGS LTD

### The Growth of Data



#### Data is growing at a 40 percent compound annual rate, reaching nearly 45 ZB by 2020



Source: Oracle, 2012

80% of Data is Unstructured

### How Much Data Are You Giving Away Now (and What Does the Future Hold?)

- Frequent Purchase Cards / Memberships
- Online shopping from Amazon and others
- Netflix
- Google
- Social Media
- FitBit /other health monitors
- Connected employee badge Humanyze
- Wearables / implants?
- Smart home applications home security, connected garage doors, doorbells, learning thermostats, house keys, home appliances, and entertainment devices
- Smartphone applications

### Lots of Primers on Big Data





### And Then There is This





### From Recent Headlines



#### Wall Street's Insatiable Lust: Data, Data, Data - wsj 9/14/2016

• The data hunter looking for meaningful data to sell to investors

#### When Information Storage Gets Under Your Skin - wsj 9/18/2016

• Radio frequency identification technology (RFID) - tiny implants can replace keys, store business cards and medical data, and eventually a lot more

# Salesforce Joins Race for Artificially Intelligent Business Software - wsj 9/18/2016

• Designed to automate tasks, predict behavior, and spotlight relevant information

#### Quants Do the Math on A New Target: Insurance - wsj 9/27/2016

• Almost instantaneous pricing and underwriting of small business policies with minimal information provided by the prospective insured

#### State Department Deploying Internet of Things Platform to Monitor Energy Use

- wsj 6/22/2016
  - Expected to manage energy use and sensor health in real time across 22,000 buildings in more than 190 countries

### Partial List of Trends



#### Use of Big Data analytics is expanding

- Data available and usage of the data continues to increase
- Predictive analytics uses data and statistical techniques to understand future trends
- Prescriptive analytics provides guidance on what to do with that future trend data example translation of risk score into actionable underwriting decision

#### Machine learning gets smarter

 Machine learning finds patterns in data and generates code to help you recognize patterns in new data; it can help create smarter applications by teaching themselves to grow and change when exposed to new data

#### Location + Big Data insights will drive mobile sales and marketing

• Real-time, targeted marketing promotions

#### **Internet of Things**

• Ability to gather and share data from everything, everywhere, is increasing

#### **Opportunities to Partner to Produce, Consume and/or Analyze Data**

### Partial List of Issues



#### **Privacy**

- Global Data Protection Regulation (GDPR) in the EU extends penalties to data owners and data processors
  - All rights for use and collection of personal information reside with the individual
  - Effective May 25,2018
- Bermuda Personal Information Protection Act (PIPA) passed in July
  - Effective data uncertain
- US expected to have some action on privacy laws in 2017

#### **Data Security**

- Imperative to properly secure the data while in use and carefully and completely dispose of the data when it is no longer needed
- Many organizations are behind on protection of data they already possess; this issue continues to grow

#### **Discriminatory Use**

- Fair Credit Reporting Act, Equal Opportunity Laws, and Federal Trade Commission Act still apply\*
- \* For more information see the Jan 2016 FTC report "Big Data A Tool for Inclusion or Exclusion"

### Why is Big Data Important to Insurers?



#### Historical Risk Assessment

- Multivariate pricing with limited variables (also more limited tools, and approved/ understood methods)
- Focus on "capturing" the right data
- Limitations due to computing power and data capture and storage costs
- Underwriting judgment key to assessing within broad categories

#### Now

- Data is everywhere it is given, purchased, and frequently just taken
- Computing power and data storage are no longer issues (data storage costs decrease as the amount of data increases)
- "Dirty" and unstructured data (text, audio, video, images) becoming easier to handle
- Expectation that "correct" price for every risk should be achievable
- Customer expectations are heightened
- Big data is enabling disruption

### Big Data is Driving InsurTech Investing



#### Big Data is creating a number of new companies that target industry disruption

- Risk assessment: Tyche Using structured and open sourced unstructured data to identify emerging risks
- Claims processes: DropIn the "Uberfication" of property damage claims adjusters
- Product design: Trov Insuring individual personal items through the phone
- Distribution models: Slice Insurance for the sharing economy
- Telematics: AssureNet Commercial telematics predicting and mitigating risks
- Emerging risks: Using structured and unstructured data to identify emerging risks
- Peer to Peer: Lemonaid "Community-based insuring economy"





TECH INVESTMENTS BY (RE)INSURERS BY STAGE



### Disruptors Target the Entire Value Chain Virtually All Data Driven/Enabled



### **Accelerators Hasten Disruption**



### PLUG AND PLAY

# Industry specific accelerators combine with industry leaders to leverage big data in pursuit of disruption

- Silicon Valley based company focusing on the creation, development, and funding of new insurance technology
  - Review 100+ startups to join the Acceleration program
  - Multi-stage vetting process for admission to the program
    - 12 week program provides mentorship, access to technology, limited capital, and corporate partners to create the ultimate startup ecosystem

#### Corporate Partners Munich RE State Farm Sun 🐼 Affac PRONTO Deloitte. TRAVELERS MAIDEN RE THRIVENT SOMPO Digital Lab TIGERRISK MS**Øamlin** τalanx. Allianz (II) **SELECTIVE**<sup>®</sup> GUY CARPENTER



### PLUG AND PLAY

The companies selected for the acceleration process range across the spectrum of analytics, technology, products, and customer engagement and all either produce, collect, or use big data

Our Startups CDriveway 🔤 🔽 sureify 💙 BIRDEYE 💭 wallflower 🛛 vidora PandaDoc 💋 LifeSaver Eafr)s CYtora Isplitsecnd (\*) safesite AMODO CLIENTDESK & Betterview Cover Slice 

The Question Is: Who is the Next Billion Dollar Company?

### How Transformative Will It Be?



### The Sexiest Job of the 21st Century: Data Analyst

Chris Morris, Special to CNBC.com Wednesday, 5 Jun 2013 | 1:00 PM ET

#### SCNBC

1710	120			Little Of	7520	00		=0.Velve	934.50
				729,00		1751.00	-6.50	+2.19%	215 55
(1880)						0.00	+1.50	+1.46%	000 50
1000					67.00	4,976,00	÷19.00	12.73%	1,882.30
SE.					1.318.00	37.09	+46.00	LD 00%	592.00
ALL .				- and set	1,728.00	218 80	0.00	1 27%	2.38
167		11533			1 072 08	96.50	+67.00 🛕	+1.37 %	17.11
- 100				384.20	1 269 00	383.70	+57.00 🔺	+1./6%	200 10
- 11			209.50	2,439,00	37.03	1.274.00	-0.72 🔻	-1.92%	Statementar
and the second	SSIS-	1994104 19400	218.30	1,106.00	218.38	1,1,100,000	-1.60 🔻	-0.73%	732.60
			95,10	2,649.00	15RM		10.10		24(6)(6)(0)
			334.20	934.50	362	2.649.00	+5.70	+1.51%	815(0)115(0)
			109.00	315.55	1.272.00		+8.00	+0.63%	70.00
			The local diversion of	597.00	2,436.00		+49.00 🔺	+2.050	10100
				2.38	2.545.00	1,902,50	A 004164	+1,05%	1,319.00
					93400	- Constant	(30×04)	12,88%	1,729.00
					31.5116		HULLON A	+1.46%	6.90

Photo: Biddlboo | Getty Images

Looking for a career change or a college major that's all but guaranteed to result in a hefty salary with copious benefits? Big data may not seem the obvious choice, but it could be your best.

### Jobs Rated Almanac Results



Career	2015 Rank	2016 Rank
Data Scientist	6	1
Statistician	4	2
Information Security Analyst	not listed	3
Audiologist	2	4
Diagnostic Medical Sonographer	not listed	5
Mathematician	3	6
Software Engineer	8	7
Computer Systems Analyst	10	8
Speech Pathologist	11	9
Actuary	1	10*

\* CAS announced the addition of a Predictive Modeling & Data Science credential in late 2016 The Institutes announced on 9/20 a new designation - Associate in Insurance Data Analytics (AIDA)

### Data Scientist: The Sexiest Job of the 21st Century But here is what may be coming ...







# New Data Sources in P&C Insurance (some are already in use)

- Auto Insurance: telematics/usage based insurance, autonomous and semiautonomous vehicles, apps/devices/in-vehicle technology identifying driver and vehicle characteristics, wearable devices/implants, social media
- Commercial Property and Homeowner Insurance: sensors/smart homes/workplaces
- CGL and Workers Compensation: wearable devices/implants, sensors/smart workplaces, sentiment analysis, connected workplace, social media
- A&H: wearables devices/implants, social media
- **Pet Insurance:** implants
- All Lines Impact: MUCH more sophistication in pricing and underwriting insurance products, new insurance products, disruption of some existing insurance products, better fraud prevention, better identification (and exploitation?) of the value of a customer, more efficient claims and litigation handling, and better/more targeted customer service

### Big Data As A Complement, Not A Replacement



- What problem do you want to solve how can you frame it?
- How can data help you to tell your story?
- How will you assess the reasonability of the data used and the conclusions drawn?
- What are the key assumptions and are they valid?
- How will you present the data to others?
- Are there ethical or fairness concerns?
- Look for the truth, not just a validation of your particular views
  - Encourage the devil's advocate

### What is the Underwriter's Role



### Ask lots of questions:

- Is the data set to be used representative of the population to be insured?
- Correlation versus causality?
  - Data scientist won't necessarily explain why correlations exist
  - What should be done with the unexpected?
- Does model address biases?
- How will the accuracy of predictions be tested over time?

#### What can users do to help data scientists?

- Describe the ideal tool/prediction/visualization/... think big (frame the problem)
- Discuss variables with the data scientist; user's domain knowledge can be extremely valuable to data scientists in deciding how variables are used/transformed/censored/interacted/filtered/...
- Meet regularly to get updates and provide feedback; have some "skin in the game" on development of data products

## Maiden's Current Focus



- Accident and Health Risk Scoring Model
  - Used internally and by clients
- Commercial Auto Risk Scoring Model
- Web Traffic Model
- Mining of Unstructured Data
  - To predict litigation trends and emerging exposures
- Prospecting Model
- Economic Impact Model



### Quants need poets and poets need quants.

The best outcomes result when they understand each others' capabilities and goals, and work collaboratively to find innovative solutions.