Big Data

Big Opportunities for Librarians 2017

Brian Hickey MICS
Senior Lecturer IS / IT

DBS Library Annual Seminar 2017
What is Big Data?

“...it describes a quantity of data that’s so large that traditional approaches to data analysis are doomed to failure. “I see big data as storytelling”

White J.M., & Cavaretta M. Data Scientists Ford Motor Company 2014
What is Big Data?

“[it’s] analysis for data that’s really messy or where you don’t know the right questions ... to make — analysis that can help you find patterns, anomalies, or new structures amidst otherwise chaotic or complex data points.”

Philip Ashlock, the chief architect of Data.gov,
What is Big Data?

“is nothing more than a tool for capturing reality—just as newspaper reporting, photography ... are. But it’s an exciting tool, because it holds the potential of capturing reality in some clearer and more accurate ways than we have been able to do in the past.”

(Leonhardt, D., New York Times)
What is Big Data?

“Big data: Endless possibilities or cradle-to-grave shackles, depending upon the political, ethical, and legal choices we make.”

(Deirdre Mulligan, Associate Professor of Cal’s School of Information)
What Happens in an Internet Minute?

1,572,877 GB of global IP data transferred

- 10 Million ads displayed
- 347,222 Tweets
- 3.3 Million pieces of content shared
- 6.9 Million messages sent

Netflix + YouTube = more than 1/2 of all traffic

- 438,801 Wiki page views
- 400 Million during Alibaba peak day sales
- 31,773 hours of music played
- 57,870 page views
- 38,194 photos uploaded
- 100 hours of video uploaded
- 138,889 hours of video watched
- 23,148 hours of video watched
- 10 Million ads displayed

And Future Growth is Staggering

2012 - 2017

By 2017, mobile traffic will have grown 13X in just 5 years

In 2017, there will be 3X more connected devices than people on Earth

All digital data created reached 4 zettabytes in 2013
Big Data = Transactions + Interactions + Observations

**WEB**
- Web logs
- Offer history
- Mobile Web
- User Click Stream

**CRM**
- ERP
  - Purchase detail
  - Purchase record
  - Payment record
- Customer Touches
- Support Contacts

**BIG DATA**
- Sensors / RFID / Devices
- User Generated Content
- Social Interactions & Feeds
- Spatial & GPS Coordinates
- External Demographics
- Business Data Feeds
- HD Video, Audio, Images
- Speech to Text
- Product/Service Logs
- SMS/MMS

**Increasing Data Variety and Complexity**

(Source: Hortonworks, Teradata, Inc., 2012)
Big Data

- Volume: Data Size
- Velocity: Speed of Change
- Variety: Different forms of Data Sources

- Value: Business Value
- Viscosity: Does it stick with you?
- Visualization: Does it tell a story at a glance?

- Veracity: Uncertainty of Data
- Variability: Inconsistencies in the data?
- Validity: Data Size

- Vulnerability: GDPR - Big Data
More on the 3 V's

**Big Data:** Expanding on 3 fronts at an increasing rate.

(Data Velocity)

(Data Volume)

(Data Variety)

(The 3 Vs that define Big Data, Soubra, D., 7/12)
Visualisation

“We hear Statistics – We Feel Stories”

(unknown, nd)

• David McArdles –
  – Information is Beautiful
    • 2013 Billion Dollar O Gram
    • Mountains out of Molehills – Media Inflated Fears
Block parties
Britain, selected manifesto promises, May 2017

Conservative
- Introduce pensions “double lock”
- Third runway at Heathrow
- More grammar schools
- More free movement of people

Liberal Democrat
- Legalise cannabis
- Free lunch at primary school
- Cap energy bills
- Nationalise rail, mail, water and buses

Labour
- Second Brexit referendum
- Reverse welfare cuts
- Protect pensions “triple lock”
- Abolish university tuition fees

UKIP
- Scrap cap on social care costs
- Drastically cut immigration
- Leave single market
- Moratorium on Muslim schools

Sources: Party manifestos; press reports
Big Data Success Stories?

- BMW
- Workday
  - Predicting when employees are going to quit their jobs
- Drug Cartels Central / South America
- Premise (Co-founder by a former Librarian)
  - Food Prices -- Early indicator of potential issues
    - Liberia
    - Argentina
- Predictive Modelling Using Big Data
  - Fraud / Customer Retention / Clinical Decision Support Systems / Child Protection / Cross Selling
Big Data Success Stories?

• PredPol
  – LAPD & Metropolitan Police
  – >60 Other Police Forces
• Facial Recognition & Physiognomy
• JP Morgan
  – Predictive algorithm knows a rogue employee before the employee even knows
• Retail
  – ApplePie
• Healthcare
  – Google & Stanford Medicine
  – Collaboration between Pharma Companies
Problems with Big Data?

• Google Flu trends

• Recent Elections

• Solid Gold Bombs
  – World War Two Slogan
  – Keep Calm And [Transitive Verb] [Direct Object]
    • “Keep Calm and ……………”
Outrage after NY paper publishes names of gun permit holders
Importance to Libraries?

Resources per Student in ARL University Libraries, 1986-2015

% Change Since 1986

-50% 0% 50% 100% 150% 200%


Interlibrary borrowing per student. (+107%)
Volumes held per student. (+54%)
Titles held per student. (+22%)
Total staff per student. (-43%)

**Monographs Purchased and Volumes Added Cross trend lines discontinued in 2011 with the advent of the revised ARL Statistics survey.
How to use this for the library

• Develop Digital Strategy
  – Log and analyse all User touch points in library
    • Book Scanning / Returns / Journals / E-books / e-audio books
    • LMS RFID Readers? / Photocopiers / Internet Usage / Others...
    • Web / Mobile
    • Tracking by Mobile / Heat Mapping

• Marketing
  – Rebrand the role of the Librarian
    • Google → a person in charge of or assisting in a library
    • Bing → A librarian is a person who works professionally in a library
    • Yahoo → A person, typically with a degree in library science, who administers or assists in a library

  – Digital Strategy
How to use this for the library

• New Revenue Streams?
  – Why no ads on Google or Facebook?
  – Identify the customer journey (Sample)

• E-discovery,
  – cataloguing and classify

• Big Data will do to multiple industries what Napster / Netflix did to Music / TV companies
Can Libraries Help?

• Information Governance Survey 2017 (UK Information Commissioner’s Office)

  Snapshot of findings –
  – 25% of Councils don’t have a Data Protection Officer – a key requirement from 06/18
  – Over 15% have no employee Data Protection training as required under GDPR
  – 33% don’t currently do privacy impact assessments on their work

“Local Authorities are complex information-handling organisations. Because they are dealing with sensitive personal information, they need to ensure that they lead the way in compliance with regulations on data and information. The ICO Survey raises some serious concerns – only 17% of Councils have an Information Asset Register, for example, which is an essential first step in compliance”

(CILIP Chair Karen McFarlane, 2017)

Use Big Data to find these gaps in the market that Library & Information Science Professionals can avail of...
Library Recommendations

• Develop a digital strategy
  – DataKind Hub
    • Data Scientist Volunteers looking to help organisations improve peoples lives
    • Digital Marketing Strategy

• Engage with the community

• Use Big Data to leverage new insights
  – Spatial relations
  – New Users / New Nationalities
The Future?
Further Reading

• *Developing Strategic Marketing Plans That Really Work*  
  *(Terry Kendrick, Facet Publishing, 2006)*

• [Dún Laoghaire-Rathdown County Library Service](#)

• [Topeka & Shawnee County Public Library](#)
Further Reading

• Carnagie UK Trust
  – Shining A light Publications on Libraries UK & Ire
  – Attitudes and usage patterns of libraries.

• Belk Library Appalachian State University
  – Stack Map
Further Reading

• Analytics On Demand for Libraries
  – Gale Cengage

Where your users are
A Final Thought…

Adapted from marketoonist.com