COMP 499 Introduction to Data Analytics

Lecture 1 — Introduction

Greg Butler

Data Science Research Centre

and

Centre for Structural and Functional Genomics

and

Computer Science and Software Engineering Concordia University, Montreal, Canada

gregb@cs.concordia.ca

Overview of Lecture

- 1. Big Data
 - Actionable Data
 - History
 - ► Five V's
 - ► Types of Jobs
 - ► Privacy and Security

Data & Feedback in Health ... Politics ... and Everything



Montreal Gazette 30 September 2016

Big Data



Actionable Data in Data-Driven (Clinical) Healthcare



(Infoway Health Canada 2016)

Big Data (http://dsrc.encs.concordia.ca/what-is-bigdata.html)

Big Data

Definition of "Big" has changed as we have become more advanced

History

Hollerith Cards 1890 (US population census)

Economic Data 1952 (GDP etc)

Computers 1959 — The First Digital Data Tsunami

World Wide Web 1990's — The Second Digital Data Tsunami

Social Media 1985 — The Third Digital Data Tsunami

Internet of Things 2000 — The Fourth Digital Data Tsunami

Big Science — 1960's onwards

Deep Knowledge — 2011 onwards

A key notion is **actionable data** that is useful in supporting decisions, determining actions, and adding value to an endeavour.

Big Data

The 5 V's

Volume: amount of data Variety: different types of data Velocity: rate at which data is generated Veracity: trustworthiness, level of noise Value: usefulness of data to a business

Drivers

Transactions Mobile Social Media Internet of Things

MGI Report

McKinsey Global Institute, *Big data: The next frontier for innovation, competition, and productivity*, May 2011.

What Happens in an Internet Minute?



(Intel 2012, http://scoop.intel.com/what-happens-in-an-internet-minute/)

Types of Jobs in Big Data

Data Analyst

Data Scientist

Data Architect

Chief Data Officer

The IT Perspective



Wang et al (Tech. Forecasting & Social Change, 2016)

The Big Data Analytics Perspective



Dimitrov (Health Informatics Research, 2016)

"**Privacy** refers to an individuals right to control the collection, use, and disclosure of his/her personal health information (PHI) and/or personal information (PI) in a manner that allows health care providers to do their work.

Security is about ensuring the information gets to the right person in a secure manner."

Ontario's Ehealth Blueprint http://www.ehealthblueprint.com

Privacy by Design 2009

Seven Foundational Principles

- 1) being proactive not reactive;
- 2) having privacy as the default setting;
- 3) having privacy embedded into design;
- 4) avoiding the pretence of false dichotomies, such as privacy vs. security;
- 5) providing full life-cycle management of data;
- 6) ensuring visibility and transparency of data; and
- 7) being user-centric

Prof. Ann Cavoukian, formerly Information and Privacy Commissioner of Ontario; now Ryerson University. http://www.privacybydesign.ca

Canada's Personal Information Protection and Electronic Documents Act (PIPEDA) 2000

Ten Privacy Principles

Accountability: An organization is responsible for personal information under its control and shall designate an individual or individuals who are accountable for the organization's compliance with the following principles. Identifying Purposes: The purposes for which personal information is collected shall be identified by the organization at or before the time the information is collected.

Consent: The knowledge and consent of the individual are required for the collection, use or disclosure of personal information, except when inappropriate.

Limiting Collection: The collection of personal information shall be limited to that which is necessary for the purposes identified by the organization. Information shall be collected by fair and lawful means.

Limiting Use, Disclosure, and Retention: Personal information shall not be used or disclosed for purposes other than those for which it was collected, except with the consent of the individual or as required by the law. Personal information shall be retained only as long as necessary for fulfilment of those purposes.

Accuracy: Personal information shall be as accurate, complete, and up-to-date as is necessary for the purposes for which it is to be used.

Safeguards: Personal information shall be protected by security safeguards appropriate to the sensitivity of the information.

Openness: An organization shall make readily available to individuals specific information about its policies and practices relating to the management of personal information.

Individual Access: Upon request, an individual shall be informed of the existence, use and disclosure of his or her personal information and shall be given access to that information. An individual shall be able to challenge the accuracy and completeness of the information and have it amended as appropriate.

Challenging Compliance: An individual shall be able to address a challenge concerning compliance with the above principles to the designated individual or individuals for the organization's compliance.

https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/

the-personal-information-protection-and-electronic-documents-act-pipeda/p_principle/

References

Infoway Health Canada: Big Data Analytics in Health White Paper, May 2013. Infoway Health Canada: Clinical Analytics in Primary Care White Paper, February 2016.

McKinsey Global Institute: Big data: The next frontier for innovation, competition, and productivity, June 2011. McKinsey: The big-data revolution in US health care: Accelerating value and innovation January 2013.

The Catalyst towards an Ontario Health Innovation Strategy, Ontario Health Innovation Council (ohic.ca), December 2014.

R. Fang et al, Computational Health Informatics in the Big Data Age: A Survey, ACM Computing Survey July 2016 http://dl.acm.org/citation.cfm?id=2932707

Y. Wang et al, Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations, Technological Forecasting and Social Change Available online 26 February 2016 http://www.sciencedirect.com/science/article/pii/S0040162516000500

D.V. Dimitrov, Medical Internet of Things and Big Data in Healthcare, Healthc Inform Res. 2016 Jul;22(3):156-163. English. http://synapse.koreamed.org/search.php?where=aview&id=10.4258/hir.2016. 22.3.156&code=1088HIR&wmode=FULL

C. Auffray et al, Making sense of big data in health research: Towards an EU action plan, Genome Medicine June 2016 http://genomemedicine.biomedcentral.com/articles/10.1186/s13073-016-0323-y

Guillaume Taglang, David B. Jackson, Use of "big data" in drug discovery and clinical trials, Gynecologic Oncology 141 (2016) 17–23. http://www.sciencedirect.com/science/article/pii/S0090825816300464

Michael KK Leung, Andrew Delong, Babak Alipanahi, Brendan J Frey, Machine Learning in Genomic Medicine: A Review of Computational Problems and Data Sets, Proceedings of the IEEE 104:1 (2016) 176–197. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7347331