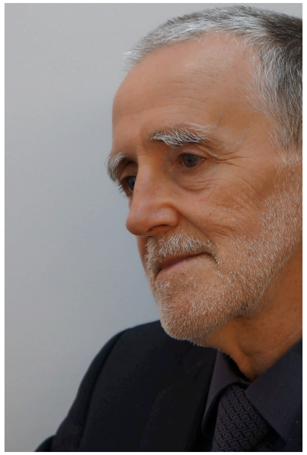
ITU Workshop on "ICT Security Standardization for Developing Countries"

(Geneva, Switzerland, 15-16 September 2014)

Information Security, PII and Big Data



Edward (Ted) Humphreys
ISO/IEC JTC 1/SC 27 (WG1 Convenor)
edwardj7@msn.com

What is BD?

- BIG DATA is high **volume**, high **velocity** and high **variety** information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making *(Gartner IT Glossary)*

- BIG DATA is a term that describes large **volumes** of high **velocity**, complex and **variable** data that require advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of the information *(TechAmerica Foundation)*

- BIG DATA is the common term used to describe the deluge of data in our networked, digitized, sensor-laden, information-driven world *(NIST Big Data Interoperability Framework: Volume 4, Security and Privacy Requirements)*

How Big is Big (volume)?

- *Kilobytes, Megabytes (large 2^{20}), Gigabytes (giant 2^{30}), Terabytes (monster 2^{40})*
- ***Petabyte*** *(2^{50}),* ***Exabyte*** *(2^{60}) – 2009 USA healthcare was 150 exabytes, FB in 2011 was 30 petabytes and in 2012 FB was growing at 500 terabytes per day*
- ***Zettabyte*** *(2^{70}) – EU estimates that about 4 ZB of data is being generated each year*
- *Yottabyte (2^{80}) - One yottabyte is approximately one septillion (10^{24}) bytes - One litre of water contains 33 Y water molecules*

How Diverse (variety)?

any type of data structured/unstructured, multiple sources, multiple formats - text, sensor data, call records, maps, audio, image, video, click streams, log files and more - hence need for Big Data Analytics

How Fast (velocity)?

fast collection/ production/processing in real time/near real time, streamed

What is BD?

- BIG DATA: BIG Potential or Big Problem
- *"The biggest advantage of big data – the ability to analyse vast quantities of data regardless of source, location or purpose – is from a legal perspective, its biggest challenge"* (Brinkman) … biggest legal problem confronting Big Data is privacy or the protection of PII (personally identifiable information)
- BIG DATA: Business Opportunity versus Risk

Who is Using BD?

- *Government*
- *Commercial sector*
- *Science, Research*
- *Education*
- *Energy Systems*
- *Healthcare Systems*
- *Transportation Systems*
- *SMART Cities*
- *Deep Learning*
- *Social Media*
- *Environmental and Ecosystems*

Information Security Risks to BD

Information security and PII of BD

- *Volume*
  - *Greater volume of data at risk (issues of multi-tiered storage and threading of data, movement, recordkeeping of gigabytes-petabytes and beyond)*
- *Variety*
  - *Risks associated with the organisation of data where there is greater degree and complexity of data from a diversity of sources etc.*
- *Velocity*
  - *Risks associated to faster production and transformation of data etc.*
- *Veracity*
  - *Magnified risks related to integrity, provenance and consistency issues etc.*
- *Volatility*
  - *Risks related to the temporal issues of data, its management, its persistent etc.*

Information Security Risks to BD

BD magnifies the concerns of information security and PII *(personally identifiable information)* and creating larger scale issues

- *Greater cyber attack surface offering the attacker a richer set of targets, multiple attack vectors …*
- *There are some aspects of BD where the traditional information security and PII methods are neither suitable, adequate nor effective and so there is a need for new and more innovative solutions need to be found*
- *The general principles relating to PII that apply to existing datasets equally apply to BD, however, BD analytics raises some new and interesting problems*

PII Preservation

- Some data subjects are ‘identifiable’ and some are ‘anonymised’
  - *Anonymization and obfuscation does not mean individuals cannot be identified: re-identification is possible either maliciously (inference attack) or otherwise*
- Data mining and BD analytics
  - *Invasion of privacy through abuse of datasets, inferencing, large scale data aggregation*
  - *Invasive marketing, consumer intelligence gathering involving PII …*
  - *Privacy, PII and the digital economy*
- PII and the Cloud
- Need for PIA (Privacy Impact Analysis)
- Legislation and Regulation on PII

PII Preservation

- PII/Privacy and the Internet/Digital Economy
  - Threat to PII/Privacy versus Threat to Business Opportunities and the Internet Economy
    - *limit business opportunity and economic growth and protect PII*
    - *allow economic growth and face legal action regarding PII*
  - Commercial “Behind the scenes” collection, exchange and analysis of customer/consumer and social media data
    - *Consumer digital media usage, social media*
    - *Family level retail transactions*
    - *Web-traffic analysis and marketing*

PII Preservation

Collection, exchange and analysis of citizen data in the field of medicine and healthcare

- *Healthcare information- collection, usage and sharing*
- *Genetic and medical research*
- *Pharmaceutical research*

Summary of Security Concerns of BD

Protective framework for

- BD that may collected and gathered from a variety of sources
  - *Covering actors - Data providers, Data owners, Data consumers, Mobile users, Social network users etc*
- BD aggregation and dissemination
  - *Data owner and data consumer contract*
- BD search and selection capability
  - *For example, protection of PII and against re-identification*
- Data management and governance
  - *Secure data storage*
  - *Attack surface reduction and attack vector reduction*
  - *Data discovery, data masking, cross-border regulation, data deletion*
- BD and PII preservation
  - *Processing steps between actors, data integrity, information assurance etc*

Information Security?

Information security for BD

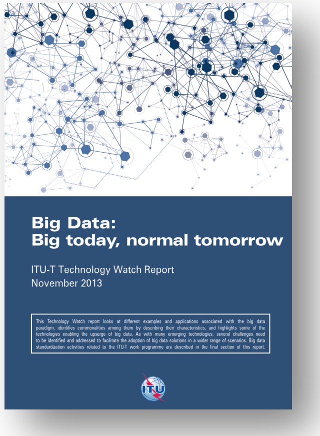
BD for information security

BD for Information Security
Real-Time Security Analytics

- BD analytics can increase the security problem but at the same time the technology can be harnessed for real-time cyber security analysis:
  - *Incident and event management (Report, analysis, evaluation …), SIEMs*
  - *Forensics*
  - *Fraud detection*
  - *IDS and IPS*
  - *National CERTs*

Examples of Current Activities

- ITU-T
  - ITU-T Technology Watch Report
  - TSAG held a BD workshop (June 2014)



Big Data:
Big today, normal tomorrow

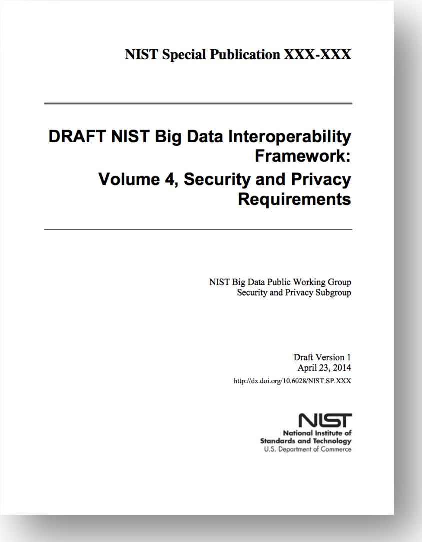
ITU-T Technology Watch Report
November 2013

This Technology Watch report looks at different examples and applications associated with the big data paradigm, identifies commonalities among them by describing their characteristics, and highlights some of the technologies enabling the upsurge of big data. As with many emerging technologies, several challenges need to be identified and addressed to facilitate the adoption of big data solutions in a wider range of scenarios. Big data standardization activities related to the ITU-T work programme are described in the final section of this report.

- ISO/IEC JTC 1
  - SG 2 – SGDB (Study Group on Big Data)
  - SC 32 – Data Management and Interchange
    - Study Group on Next Generation Analytics
    - Link to SC6 and SC 39
  - Other interested groups
    - *SC 27 (IT security)*
    - *SC 32 (Document Description and Processing Languages)*
    - *SC 38 (Distributed Application Platforms and Services (DAPS)) – Web services, SOA and Cloud*

Examples of Current Activities

- NIST
  - NBD-WG (http://bigdatawg.nist.gov)
- IEEE
  - BigData 2014, Cloud
  - Computational intelligence and BD
  - Data analytics for BD security
- CSA, OASIS



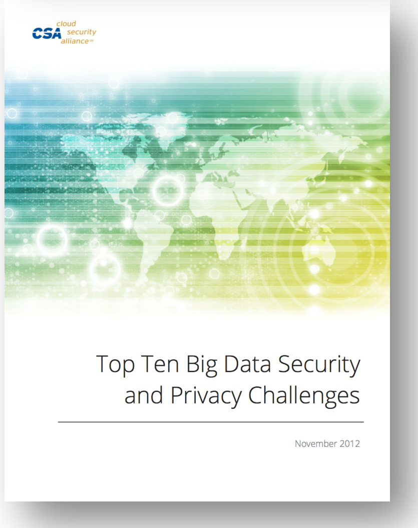
NIST Special Publication XXX-XXX

DRAFT NIST Big Data Interoperability Framework:
Volume 4, Security and Privacy Requirements

NIST Big Data Public Working Group
Security and Privacy Subgroup

Draft Version 1
April 23, 2014
http://dx.doi.org/10.6028/NIST.SP.XXX

NIST
National Institute of
Standards and Technology
U.S. Department of Commerce



CSA cloud security alliance

Top Ten Big Data Security
and Privacy Challenges

November 2012

Future Activities Needed

More work needed on infrastructure security, data privacy (PII), GRC, data management and integrity and reactive security

- *Research*
- *Standards*
- *Regulation*

Thanks for Listening

Edward (Ted) Humphreys

edwardj7@msn.com