EMIRATES
IDENTITY
AUTHORITY

Partners in Building UAE's Security & Economy

Privacy in the Age of Big Data

Presented by Dr. Ali M. Al-Khouri

In: Big Data Systems, Applications and Privacy Conference – Organised by New York University, Abu Dhabi, UAE.
10 – 11 March 2013 | Injazat Data Systems | Abu Dhabi | UAE

Federal Authority | هيئـــة اتحـــادية
Our Vision: To be a role model and reference point in proofing individual identity and build wealth informatics
that guarantees innovative and sophisticated services for the benefit of UAE
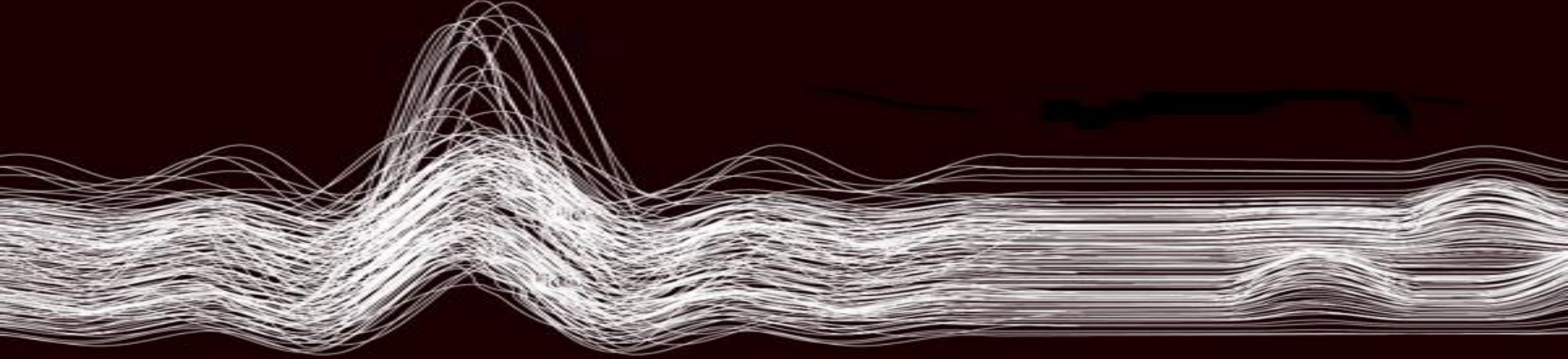
Agenda

- Introduction: The Age of Big Data

- Big Data Characteristics

- Constructing Identity from Digital Behaviour

- Government IdM and Privacy

- Concluding Remarks

Agenda

- Introduction: The Age of Big Data

- Big Data Characteristics

- Constructing Identity from Digital Behaviour

- Government IdM and Privacy

- Concluding Remarks

Big Data in Digital World ?

- The **proliferation of modern technologies**, smart devices and popularity of social networking is generating unprecedented amounts of data, both **structured and unstructured**, whether it be **text, audio or video.**

- Data has become **ubiquitous** in this digital world.

- Sheer **amount of bytes** that we generate daily is indeed **mind boggling**!

The Age of Big Data ?
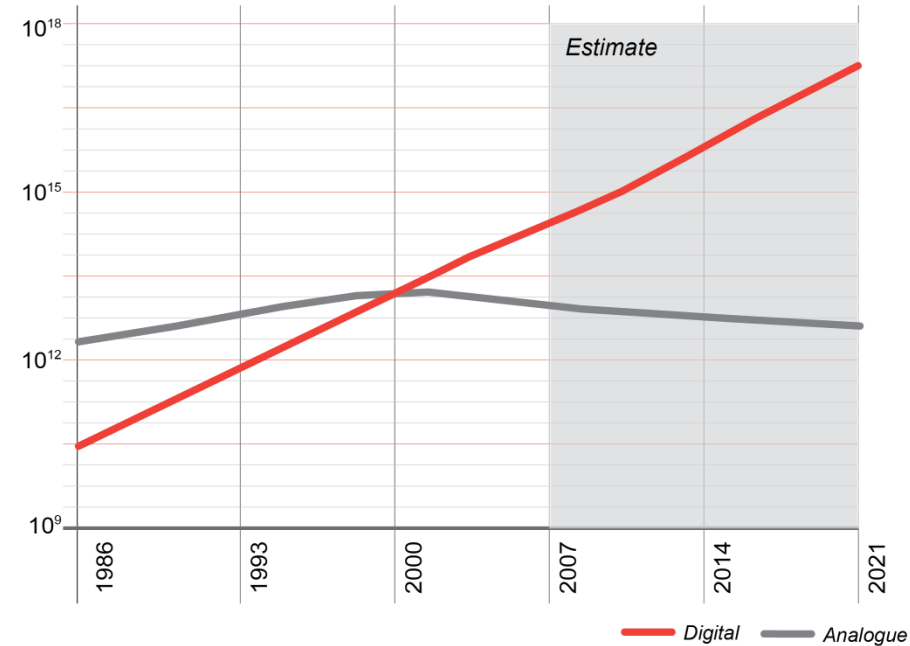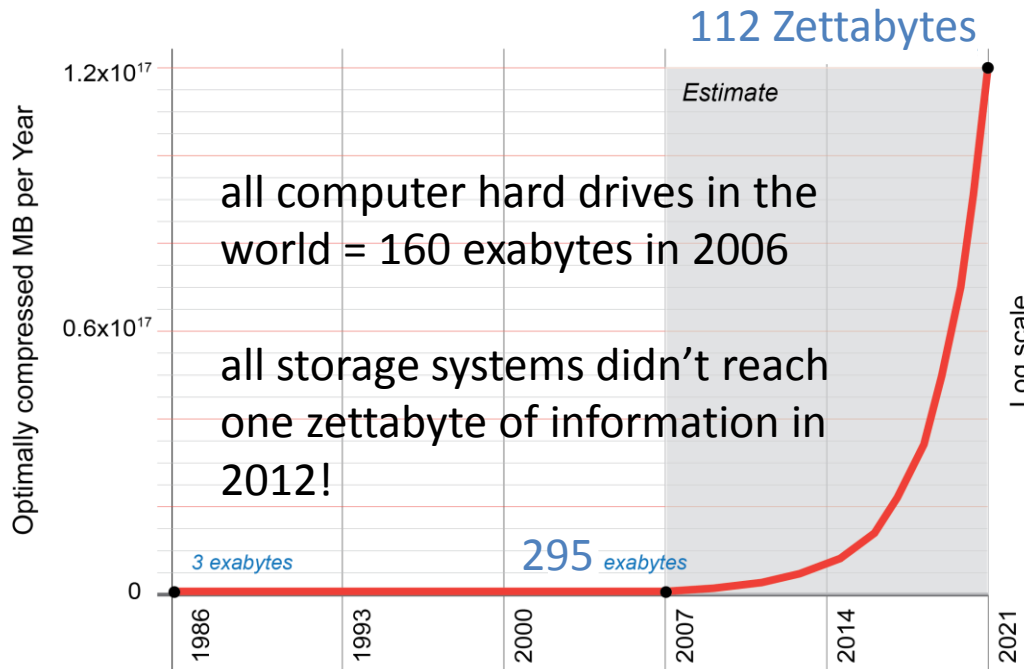
every day, we create
2.5 quintillion bytes
of data ..

90% of the data in the
world today has been created
in the **last two years**.

IBM Report

We are truly in a
digital explosion era!

This data is
big data

112 Zettabytes

all computer hard drives in the world = 160 exabytes in 2006

all storage systems didn't reach one zettabyte of information in 2012!

3 exabytes   **295** *exabytes*

Estimate

Optimally compressed MB per Year

Log scale

Digital   *Analogue*

Global Growth of Digital Storage Capacity 1986-2021

By 2020, 112 Zettabytes of data (75% growth rate)

One Zettabyte = 1,000,000,000,000,000,000,000 byte, or 1000 exabytes

Big Data ?

- is **generated from** practically **every where**:

  - Social media sites (facebook, twitter, linked–in..)
  - digital pictures and videos,
  - e–mails,
  - purchase transaction records,
  - cell phone, GPS signals,
  - geo–stationary satellites,
  - meteorological sensors
  - to name a few.

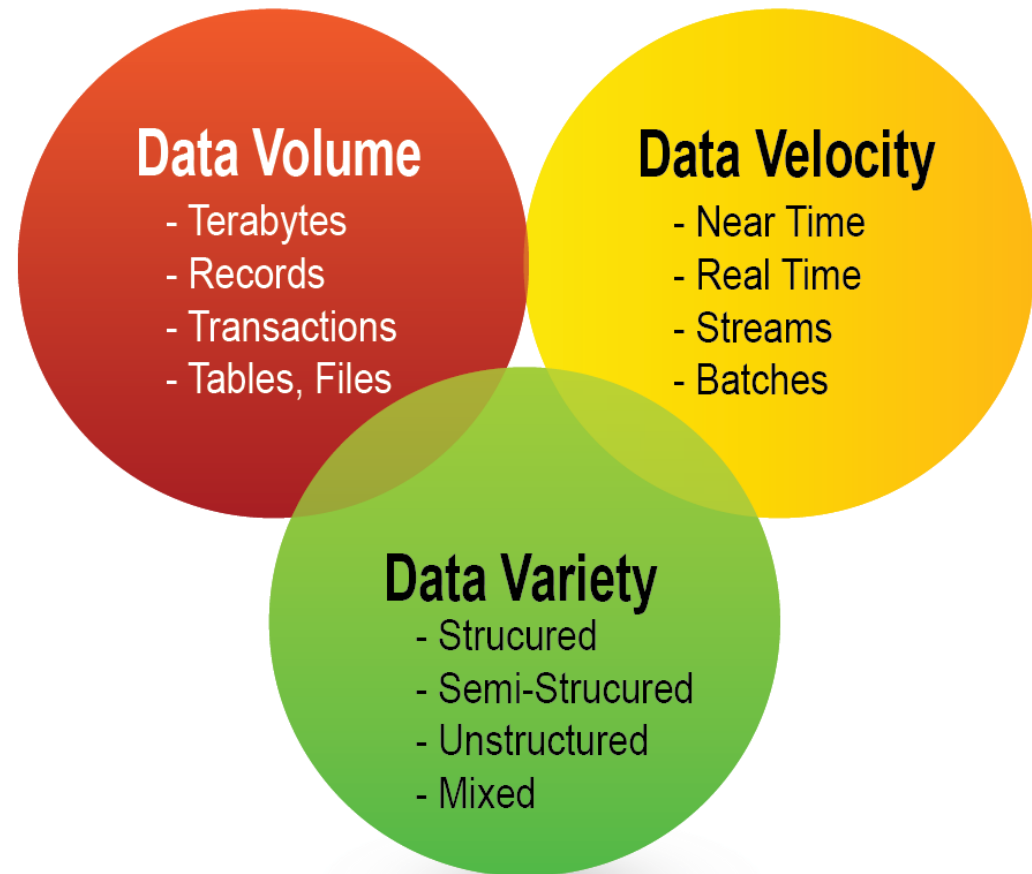- This **continues to grow** at an **exponential** rate.

Agenda

- Introduction: The Age of Big Data
- Big Data Characteristics
- Constructing Identity from Digital Behaviour
- Government IdM and Privacy
- Concluding Remarks

Big Data characteristics?

United Arab Emirates

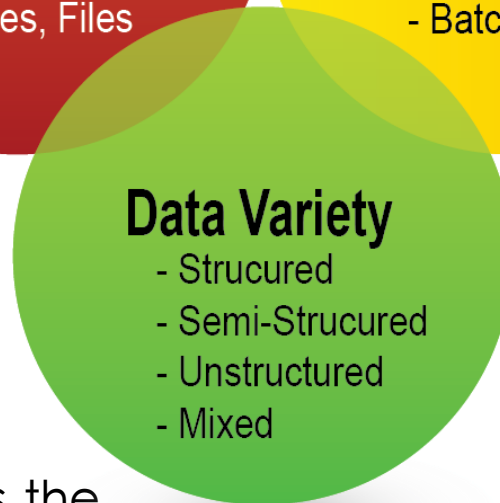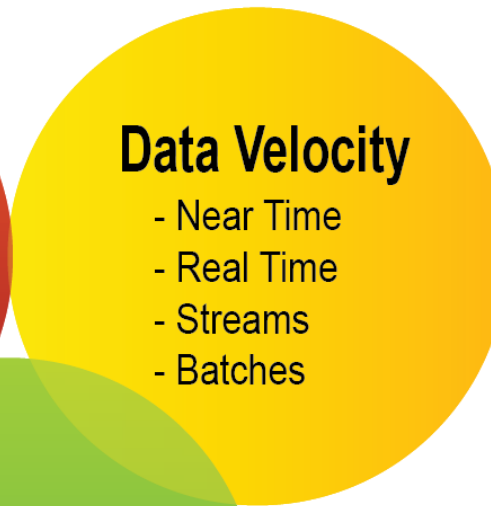هيئــة
الإمــارات
للهويــة
EMIRATES
IDENTITY
AUTHORITY

- Big Data has come to be **characterized** by the **V**olume, **V**elocity and **V**ariety of Data that is generated.

- These constitute the **3Vs** of Big Data.

Data Volume
- Terabytes
- Records
- Transactions
- Tables, Files

Data Velocity
- Near Time
- Real Time
- Streams
- Batches

Data Variety
- Strucured
- Semi-Strucured
- Unstructured
- Mixed

TWDI Best Practices Report Q4 2011- Big Data Analysis- Philip Russom
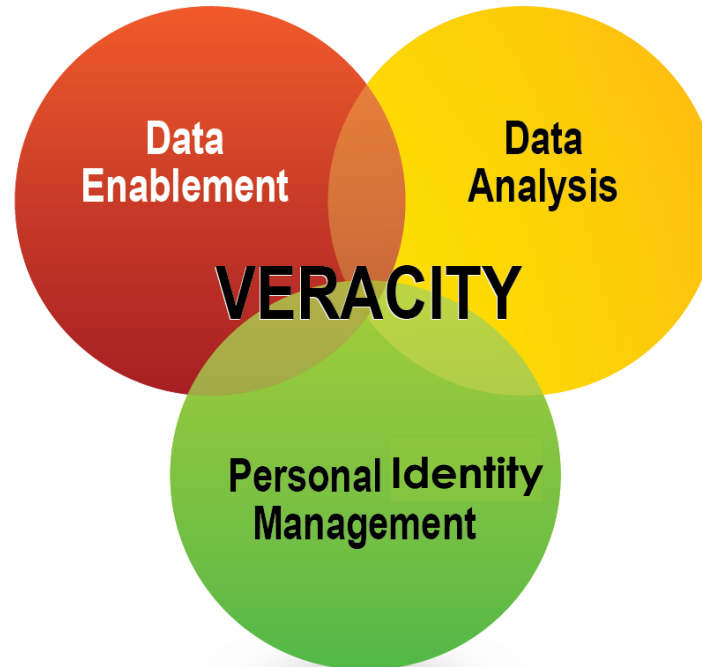
4th 'V' of Big Data ?

- However, there is the fourth dimension – a fourth V– **Veracity** which **encompasses the 3Vs**!

- Veracity provides the **confidence** in **the truthfulness of the data.**

Data Volume
- Terabytes
- Records
- Transactions
- Tables, Fil

Data Velocity
- Near Time
- Real Time
- Streams
- Batches

VERACITY

Data Variety
- Strucured
- Semi-Strucured
- Unstructured
- Mixed

TWDI Best Practices Report Q4 2011- Big Data Analysis- Philip Russom
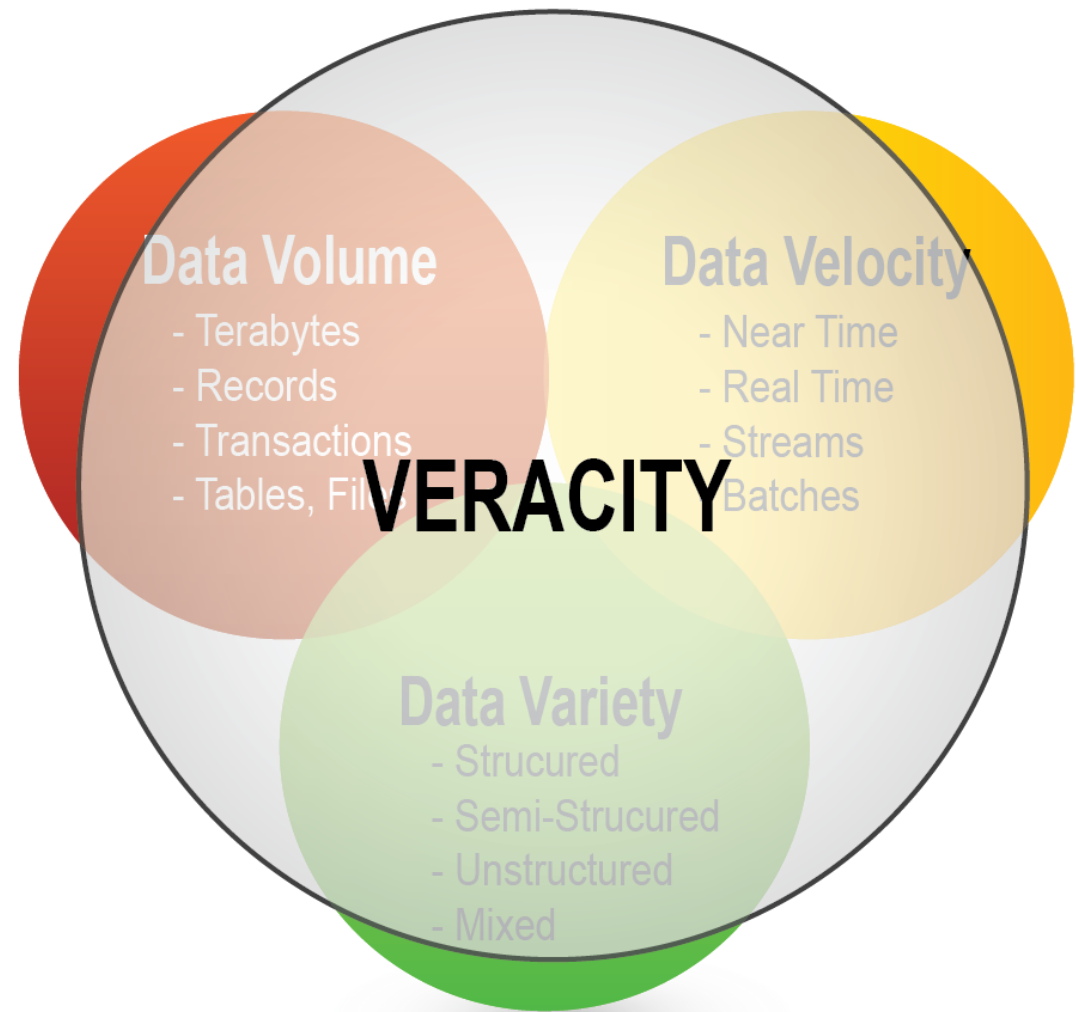
Personal ID and Big Data ?

Veracity of Data itself can be depicted using three dimensions:

1. Veracity of Data is established by **how the data itself is enabled**– which stands for the **source of data**.

Data Enablement

Data Analysis

VERACITY

Personal Identity Management

2. Veracity of Data is established by the means and **methods of analysis**– **providing discernible information**

3. Veracity of Data is then **characterized by Personal Identity Management**. This is the **critical dimension** of **Big Data Veracity**.
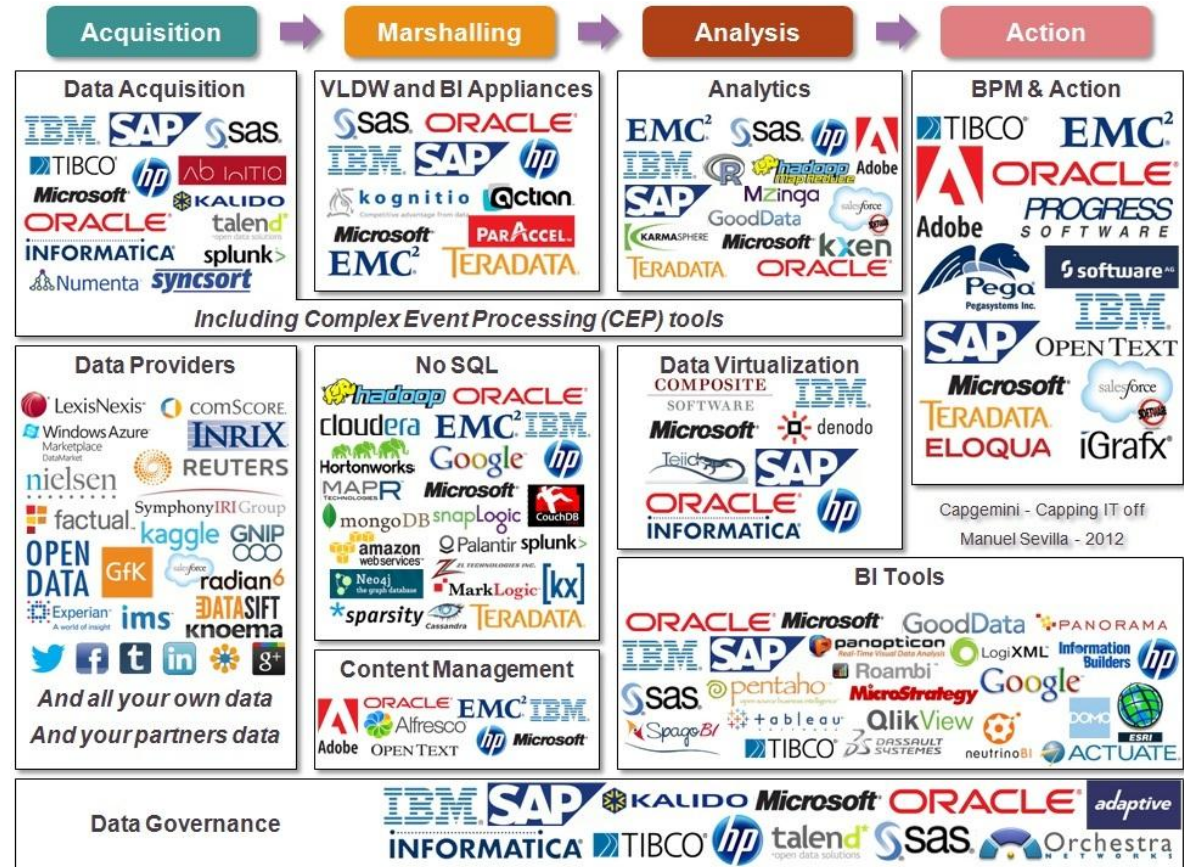
Big Data ?

- is **not a new or unknown phenomenon**.

- Big Data as **Volume Data** has been used in **Clinical Trials** since long, resulting in many a **ground breaking medicines**.

- Big Data as Volume Data has been in existence in **DNA mapping and Genetics** leading to many life–saving healthcare procedures.

- While the **Healthcare industry** has been the **initiator of big data analysis**, retailers and marketing organizations have now started to make use of Big Data to further their **commercial activities**.

- **Numerous web analytics companies** have now made it their **business to collect** all this trail and **construct meaningful data** from this huge data.

- Billions of posts in social networks, blogs, commerce sites, e-mails, text messages, utility payments are being **"piggy-backed"** to **result in patterns** of the **digital interactions and individual behavior patterns** then **constructed** from here.

- This is brilliant from a **marketing point of view**!

Capgemini - Capping IT off
Manuel Sevilla - 2012

But, from a **privacy perspective**, this could be construed as a blatant **violation of individual privacy**.

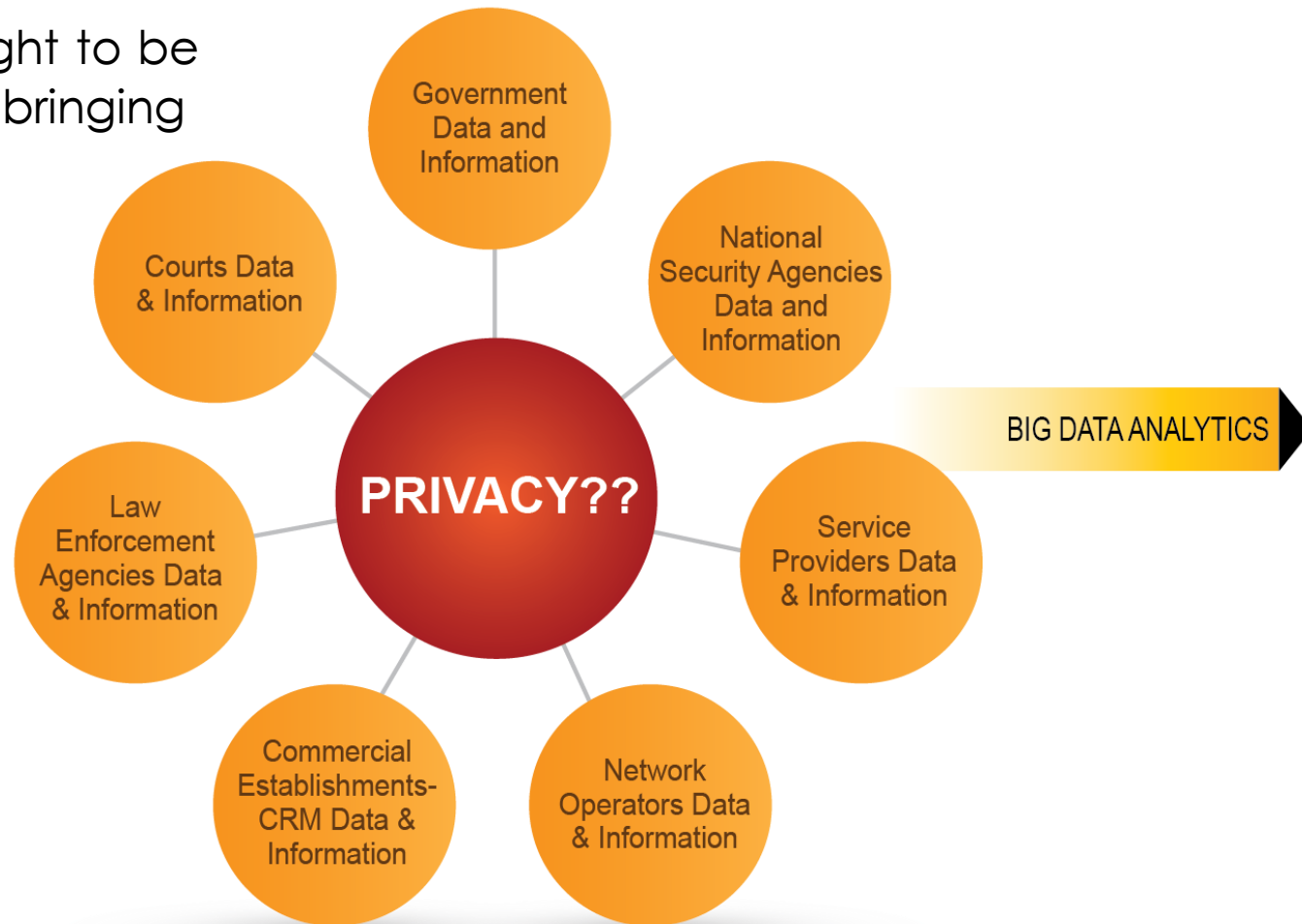# Agenda

- Introduction: The Age of Big Data

- Big Data Characteristics

- Constructing Identity from Digital Behaviour

- Government IdM and Privacy

- Concluding Remarks

Personal ID and Digital Behaviour ?

- as **individuals,** we **Interact**, we **Transact**, we **Collaborate**, we **Communicate** on a daily basis

- .. which is sought to be **analyzed** for bringing commercial entities closer to the psyche of the individual (knowing customers – personalising services and products).

Government Data and Information

National Security Agencies Data and Information

Courts Data & Information

PRIVACY??

Service Providers Data & Information

Law Enforcement Agencies Data & Information

Commercial Establishments- CRM Data & Information

Network Operators Data & Information

BIG DATA ANALYTICS

DIGITAL BEHAVIOUR

Big Data as in **information and datasets** have started to be cultivated based on the **digital data** we leave in **our transactions**.

# Digital Behaviour and Data Sets ?

- We live in the **digital world** and are **ubiquitously connected**.

- The **Phone** is no longer a <u>calling device</u>.

- Our laptops, Tablets, Phablets and Phones keep us **constantly engaged**.

- **Every interaction** leaves a **huge trail of data** leaving behind **bits and pieces** of us <u>with pointers</u> to our real behaviour.

- This **raises** then many **issues governing** **privacy**.
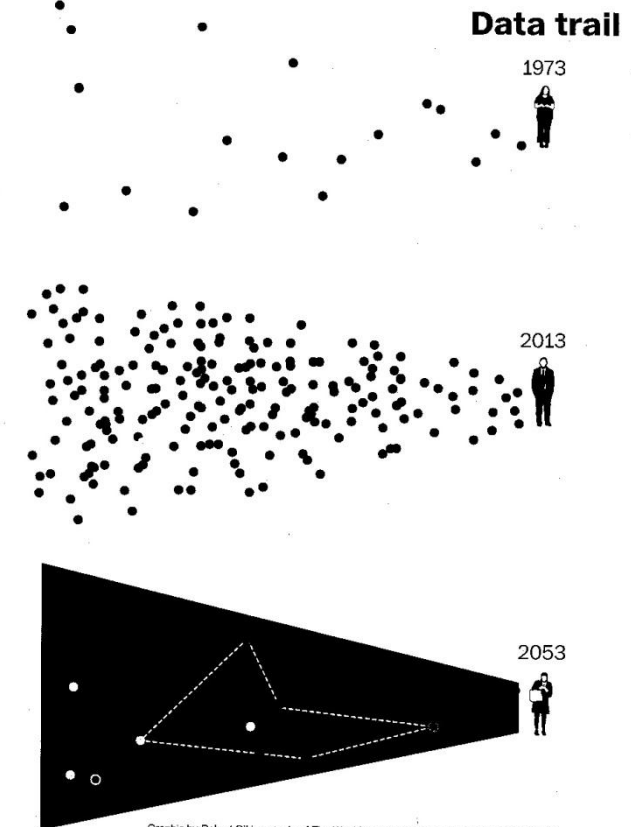
Digital Behaviour and Data Sets ?

- It is understood that **collection of information** from **digital transactions and interactions** is something that is **unstoppable**.

- Whether we like it or not, the **amount of digital trail** we leave behind in the **e-world** is **amazingly large**.

- This **digital trail** when **analyzed** is almost like **a signature that we leave behind**, making it very **easy for analysts to identify us as individuals** in the purported anonymity of the web.
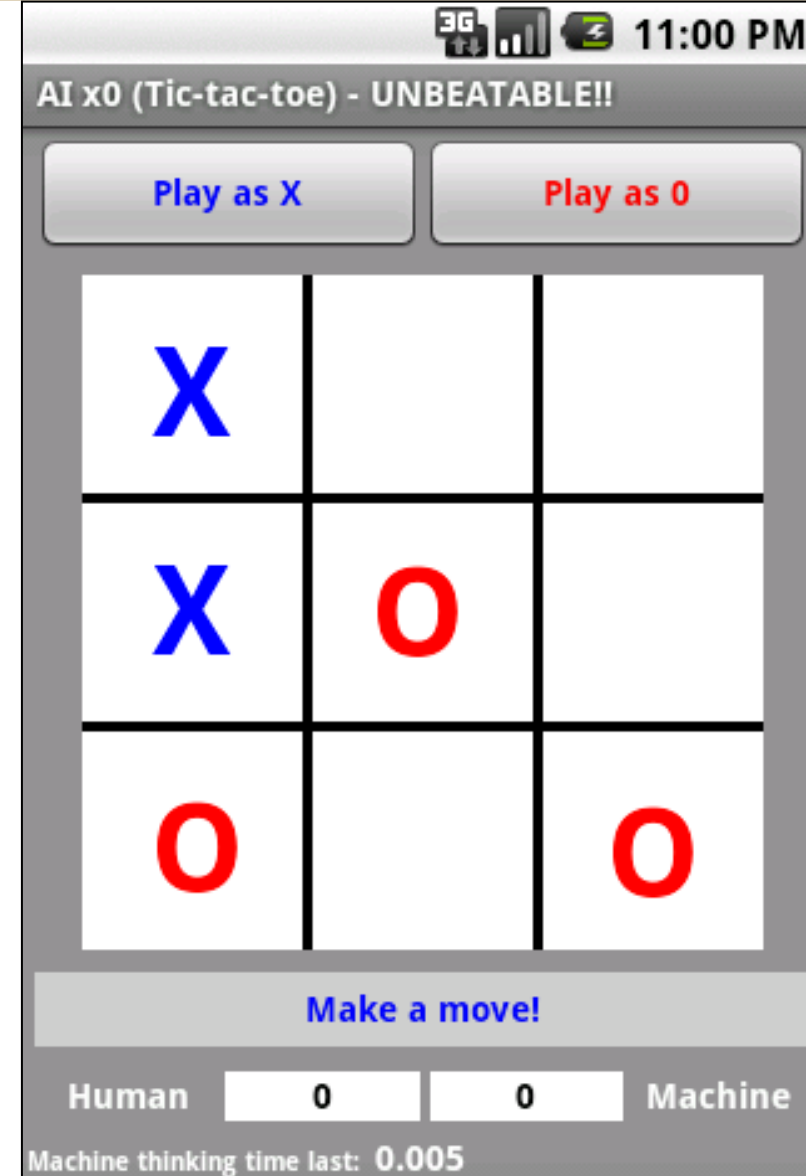
Anonymity ?

- The West, particularly the **USA and the EU** have made **conscious moves** to **protect individual privacy** from being abused with **legal provisions**.

- **Anonymity** has been the key consideration on which the legal provisions have been made so far.

- However, it has been proved beyond any reasonable doubt that **anonymity is not guaranteed** even when **personal identifiers are removed from the data sets for analysis**.

- Even **IP Addresses** that relate **to machines** have been cracked to relate them **to personal users**.

- So, **where does this leave an individual with respect to his/ her privacy?**

Data trail

1973

2013

2053

Graphic by Robert O'Harrow, Jr. of The Washington Post and author of No Place to Hide.

Data can be either useful or perfectly anonymous but never both.

PAUL OHM

Associate Professor of Law
University of Colorado Law School

How True?!

.. while **anonymized information** <u>will always carry some risk of re-identification</u> . . . . many of the most pressing <u>privacy risks</u> . . . exist only if there is certainty in re-identification, that is **if the information can be authenticated**. As uncertainty is introduced into the re-identification equation, we **cannot know that the information truly corresponds to a particular individual**; it becomes **more anonymous as larger amounts of uncertainty are introduced**.

Betsy Masiello
Global Public Policy at Google

Alma Whitten
Privacy for Product and Engineering at Google

Betsy Masiello, Alma Whitten: Engineering Privacy in an Age of Information Abundance. AAAI Spring Symposium: Intelligent Information Privacy Management 2010

Agenda

- Introduction: The Age of Big Data

- Big Data Characteristics

- Constructing Identity from Digital Behaviour

- Government IdM and Privacy

- Concluding Remarks

National ID for Big Data Analysis

- an intense role of **Government Issued Personal Identity** as a **Unique** National ID.

- this will be **extremely critical** in Big Data and **Big Data Analytics** providing the required **privacy in anonymity and** yet provide **meaningful data for analysis**.
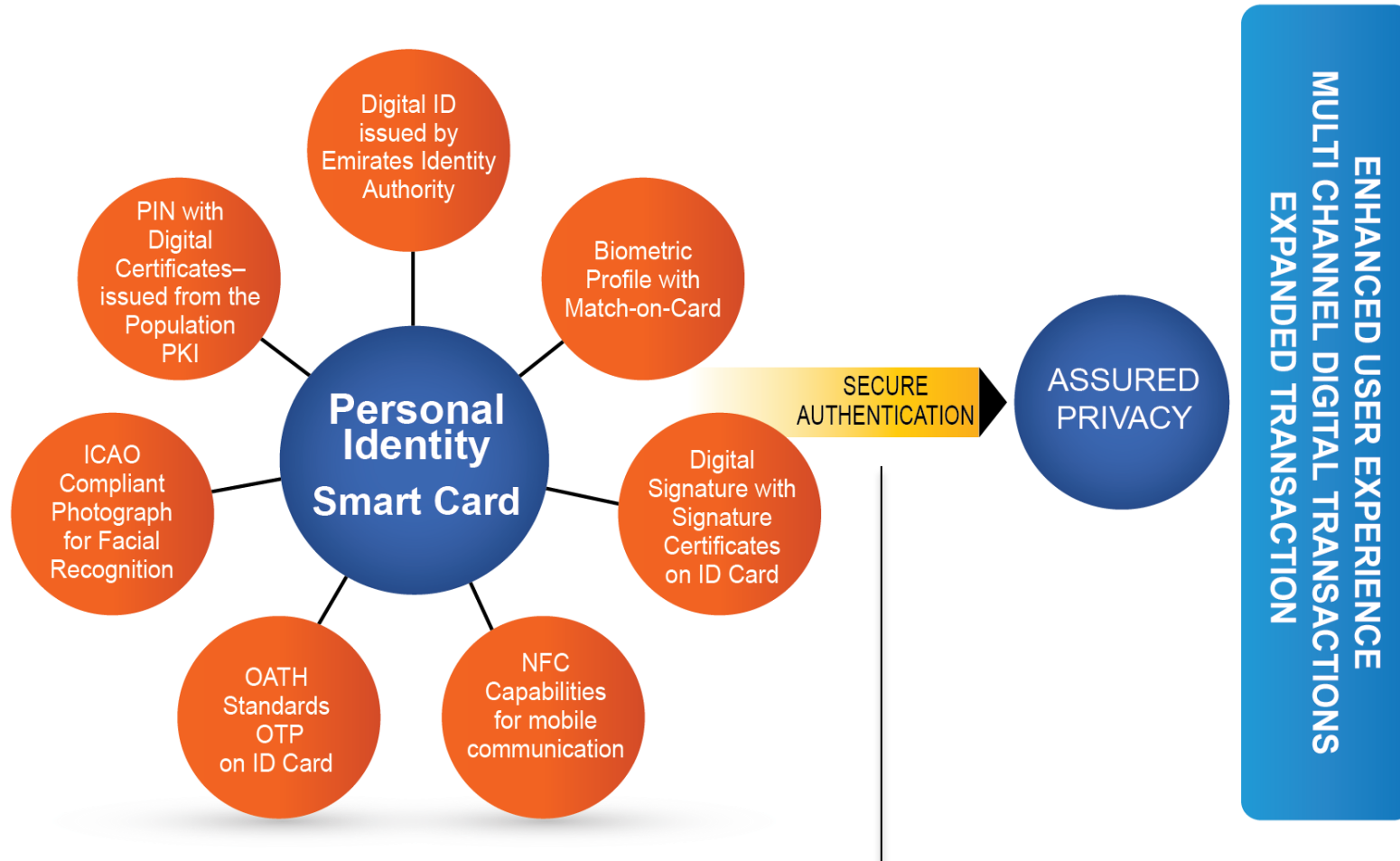
National ID for Big Data Analysis

- A **secure ID** would **encourage users** to be **engaged more actively** and more expansively in the **digital world**.

- It is in this context that the **Digital Identity profile** provided and packaged in a **Secure Smart Card** is expected to **play a pivotal role**.
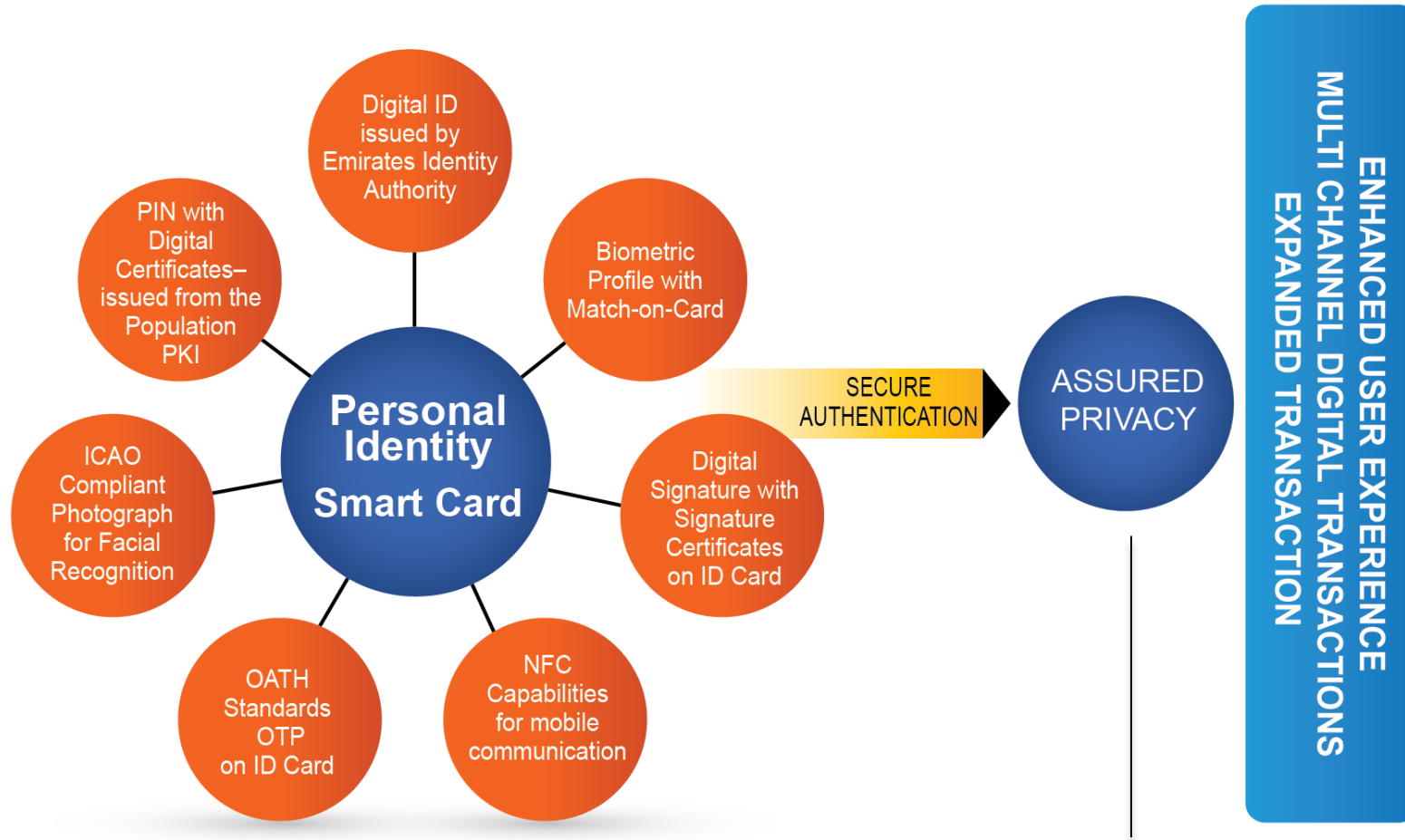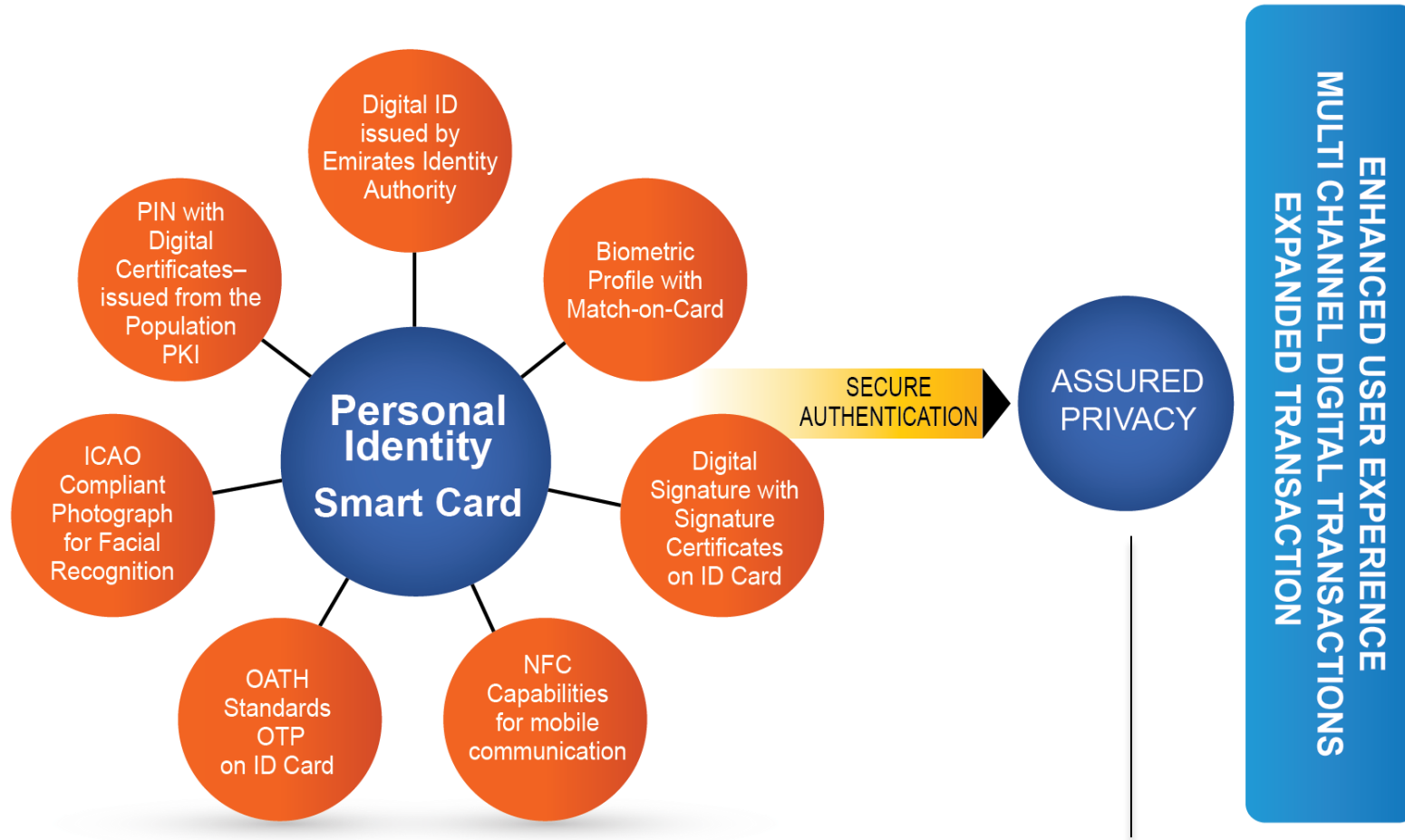
Personal ID and UAE National ID

Digital ID issued by Emirates Identity Authority

PIN with Digital Certificates– issued from the Population PKI

Biometric Profile with Match-on-Card

ICAO Compliant Photograph for Facial Recognition

Personal Identity Smart Card

Digital Signature with Signature Certificates on ID Card

OATH Standards OTP on ID Card

NFC Capabilities for mobile communication

SECURE AUTHENTICATION

ASSURED PRIVACY

ENHANCED USER EXPERIENCE
MULTI CHANNEL DIGITAL TRANSACTIONS
EXPANDED TRANSACTION

The **Digital Identity** provides the perfect **PROXY** for the personal identity.
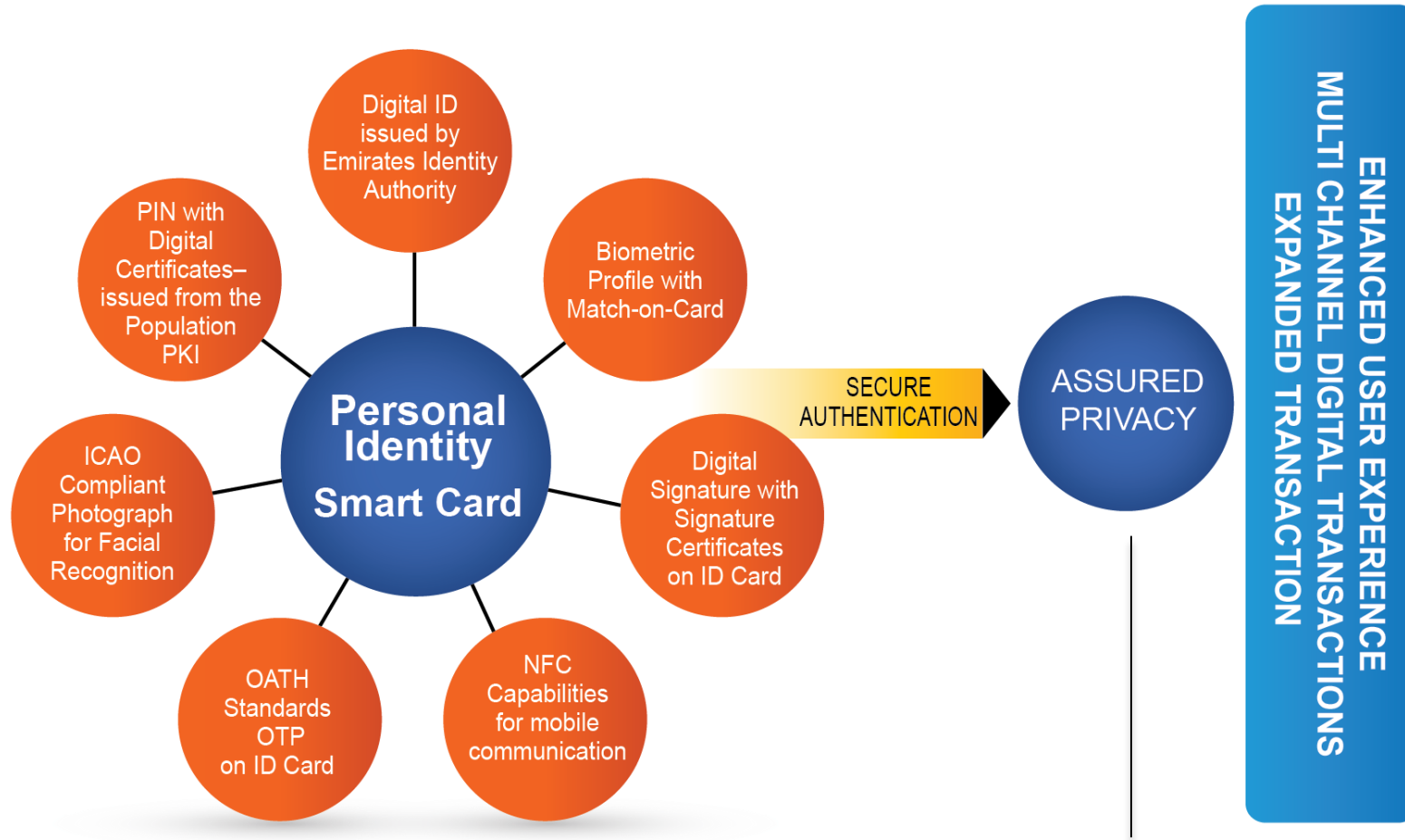
Personal ID and UAE National ID

The **individuals** would be **known and authenticated** as genuine persons **by the** National Identity Authority providing a THIRD PARTY authentication for the identity.

Personal ID and UAE National ID

Thus, **service providers** can **identify the potential service seeker securely from the authentication** provided by the Emirates ID Authority.
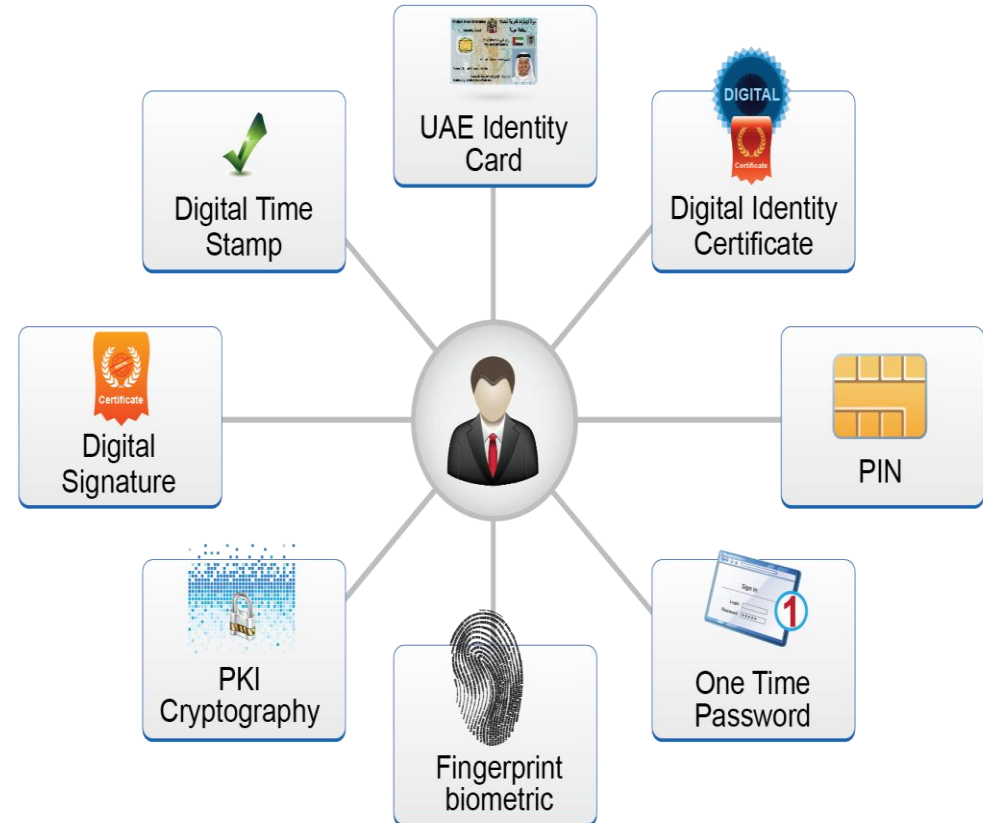
Personal ID and UAE National ID

An **individual** will then be able to **transact and interact** freely **without compromising his/her personal identity**.

National ID for Big Data Analysis

- The UAE National ID Card is designed to provide **multi factor authentication**.

- At the core of the Digital ID is the **Unique Identity Number** issued by the Emirates ID Authority.

- With this ID is associated with the **Digital Identity Profile** consisting of **Biometric Data–Fingerprints and ICAO compliant photograph**, and **Digital Certificates** issued for **Identification and Signatures** issued as a **Secure Smart Card.**

Personal ID and UAE National ID

- Apart from the **classical ID verification and authentication methods**, projects are underway to upgrade the features to provide **OTP** (One–time–password) as per OATH Standards.

- The ID card is a **combi–card** to work with **contactless readers**.

- Projects are underway to develop the **NFC** **enabled authentication methods** using the **GSM mobile phones and other mobile devices**.

- Each of the **Digital ID profile components** provides the **ability for the user to be verified and further authenticated** for access.

United Arab Emirates ... دولة الإمارات العربية المتحدة
Identity Card ... بطاقة هوية
ID Number / رقم الهوية
784-2002-8550305-0
الإسم: أحمد محمد عبد الله
Name: Ahmed Mohamed Abdulla
الجنسية: الإمارات العربية المتحدة
Nationality: United Arab Emirates

# Protection of Individual Privacy

- The Emirates ID Authority provides the necessary **credential verification on the web**.

- The **National ID Card** could be used only with the **Digital Credentials on the web transactions**.

- Emirates ID Authority **does not share the personal information but provides only credential verification**
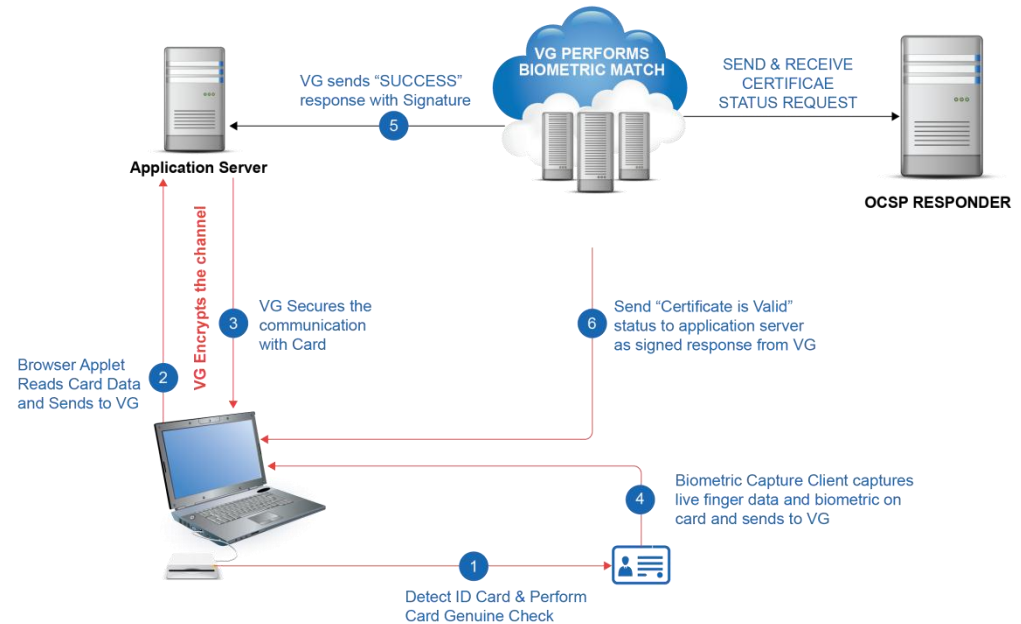
Protection of Individual Privacy

- **Service Providers** are accorded with **Verification and Authentication Services** enabling secure remote transactions.

- **Service Seekers** remain anonymous on the web since only **Digital Certificates or Biometrics** would be used to establish credential verification.
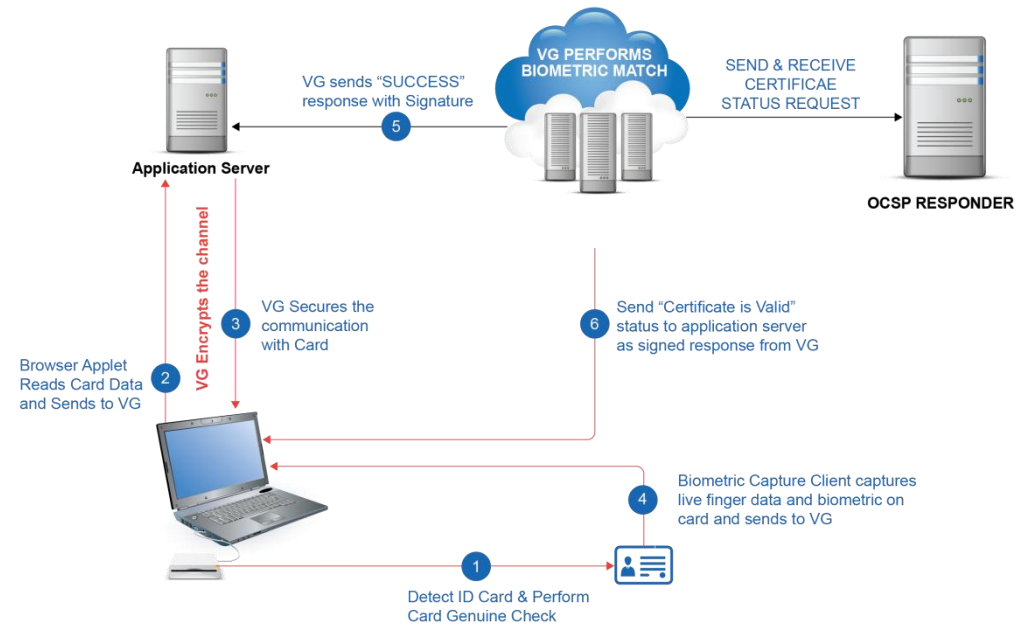
Agenda

- Introduction: The Age of Big Data

- Big Data Characteristics

- Constructing Identity from Digital Behaviour

- Government IdM and Privacy

- Concluding Remarks

- Big Data should be viewed to be about **building Knowledge** to support social, environmental, and economical development.

- i.e., to **improve efficiency, speed & accuracy of decisions**, ability to **forecast**, cut **costs**, save **energy**, improve **services**, optimise **infrastructure**, enhance citizens **quality of life**, reduce environmental footprints, fuel **innovation** and diverse **sustainable economic growth**.
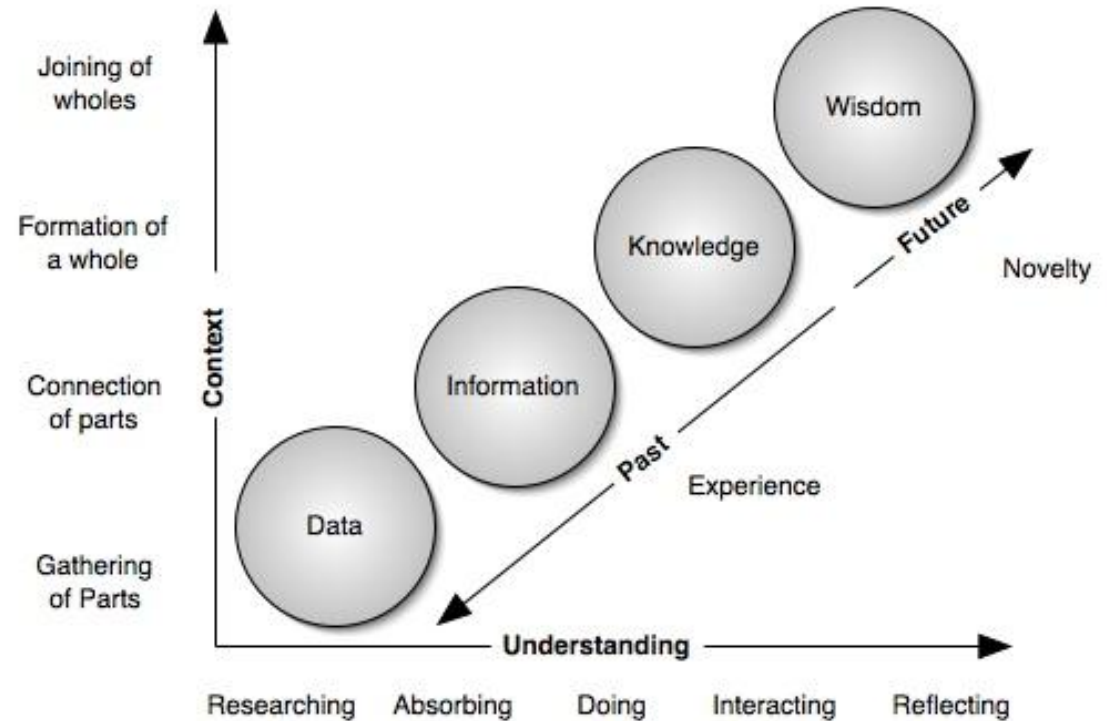
- **Complexity** will remain an issue!

- Successfully **exploiting the value** in big data requires **experimentation and exploration**.

- **Private sector** will still lead the game, as Big Data will be viewed as a **source of competition and growth**.

- **Public sector** will need to take Big Data **more seriously**, and put in place data–strategies to create new waves of productivity growth.

- **Main challenge:** shortage of **skills** [USA by 2018 will face shortage of around 2 million managers and analysts with know–how to create and use big data to make effective decisions].

- Need to **address policies related to privacy, security**, etc.

UAE National ID and Privacy

- **All data should be treated as personally identifiable** and subjected to the **regulatory framework**.

- UAE (Emirates ID Authority) would provide this **regulatory framework to ensure identity protection**.

- **Technical and Technology Solutions** are already in place to **support the regulatory and legal framework**.

More Information:

Read our recent research from:
http://www.emiratesid.gov.ae/ar/media-center/publications.aspx
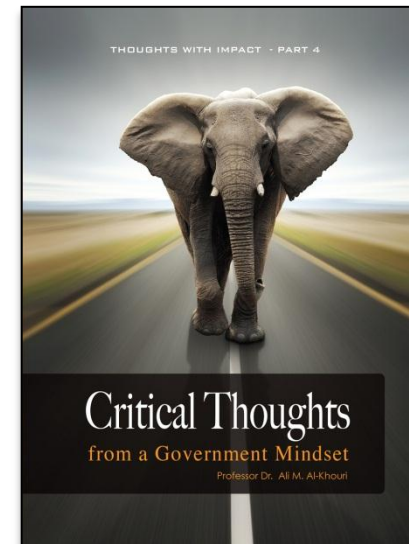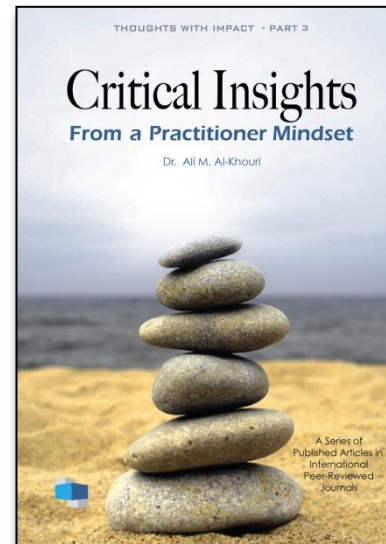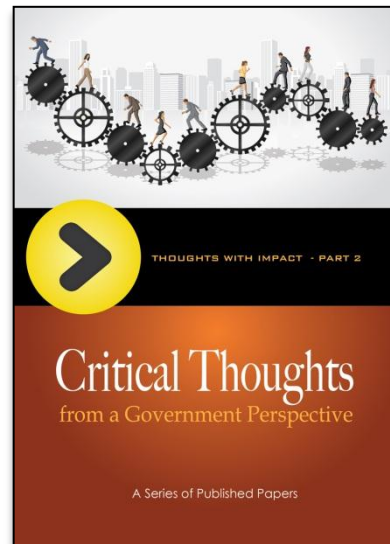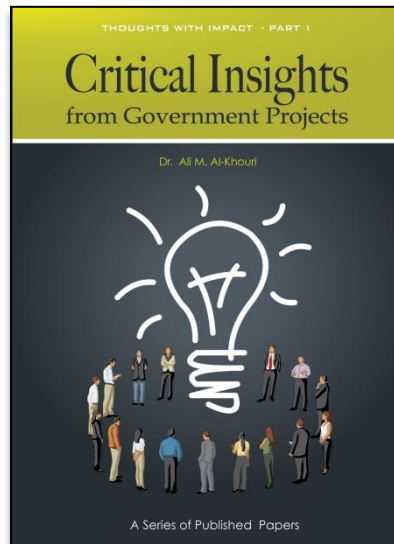
You can use all the quantitative data you can get, but you still have to **distrust it** and **use your own intelligence and judgment**.

[Alvin Toffler]

Dr. Ail M. Al-Khouri
Director General |
Emirates Identity Authority |
United Arab Emirates |
ali.alkhouri@emiratesid.ae |
www.emiratesid.ae

Thank you