# Navigating Your GRC Journey

## Convergence of "Big Data" and Data Governance, Risk & Compliance

Gareth Martin
*Big Data Analytics Lead EMEA, HP*

Grace Crossley
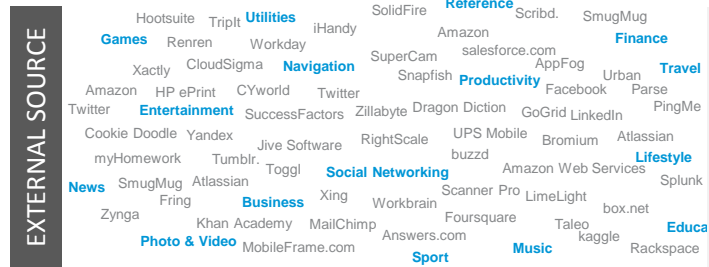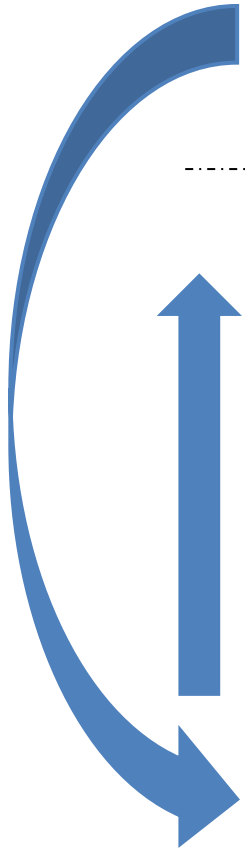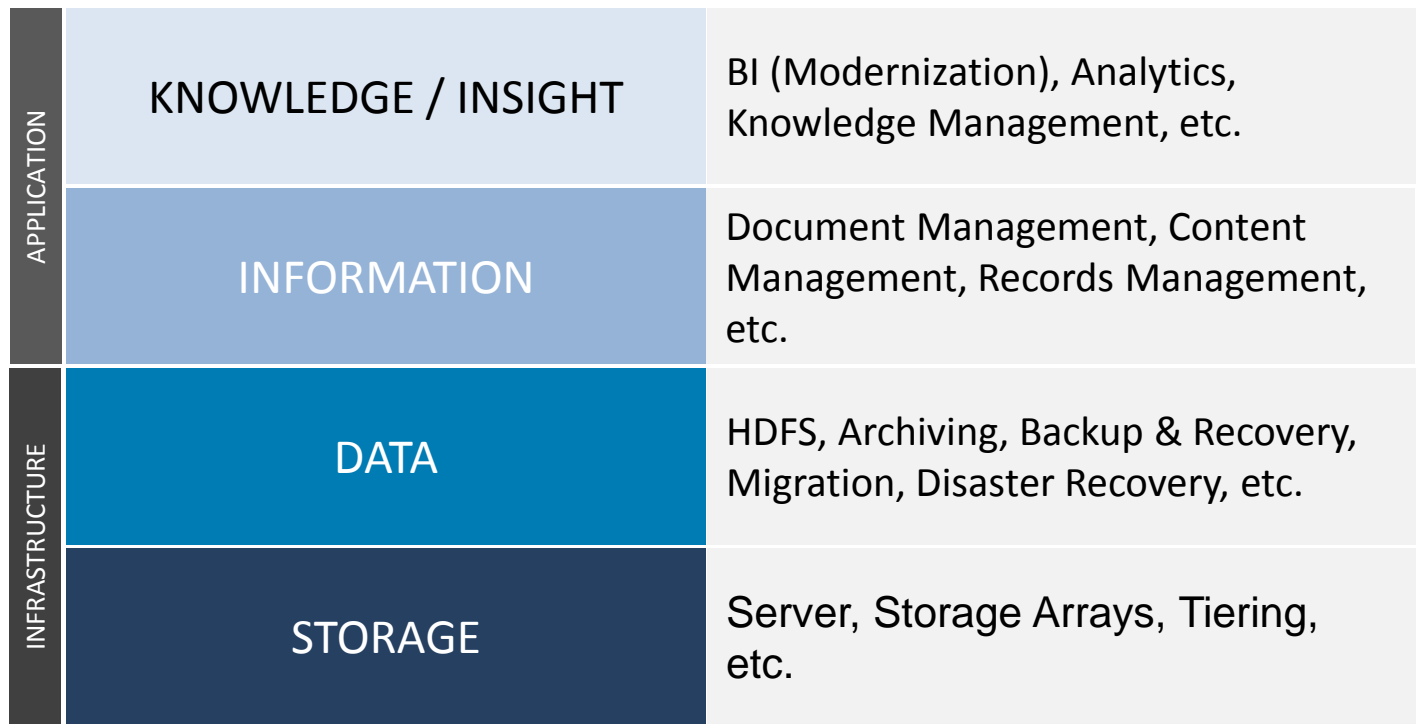*Information Governance Lead EMEA, HP*

# AGENDA

- **INTRODUCTION**

- **EXAMPLES OF GRC IN BIG DATA**

- **BREAKING IT DOWN – OUR FRAMEWORK**

- **QUESTIONS**

# Introduction

BIG DATA

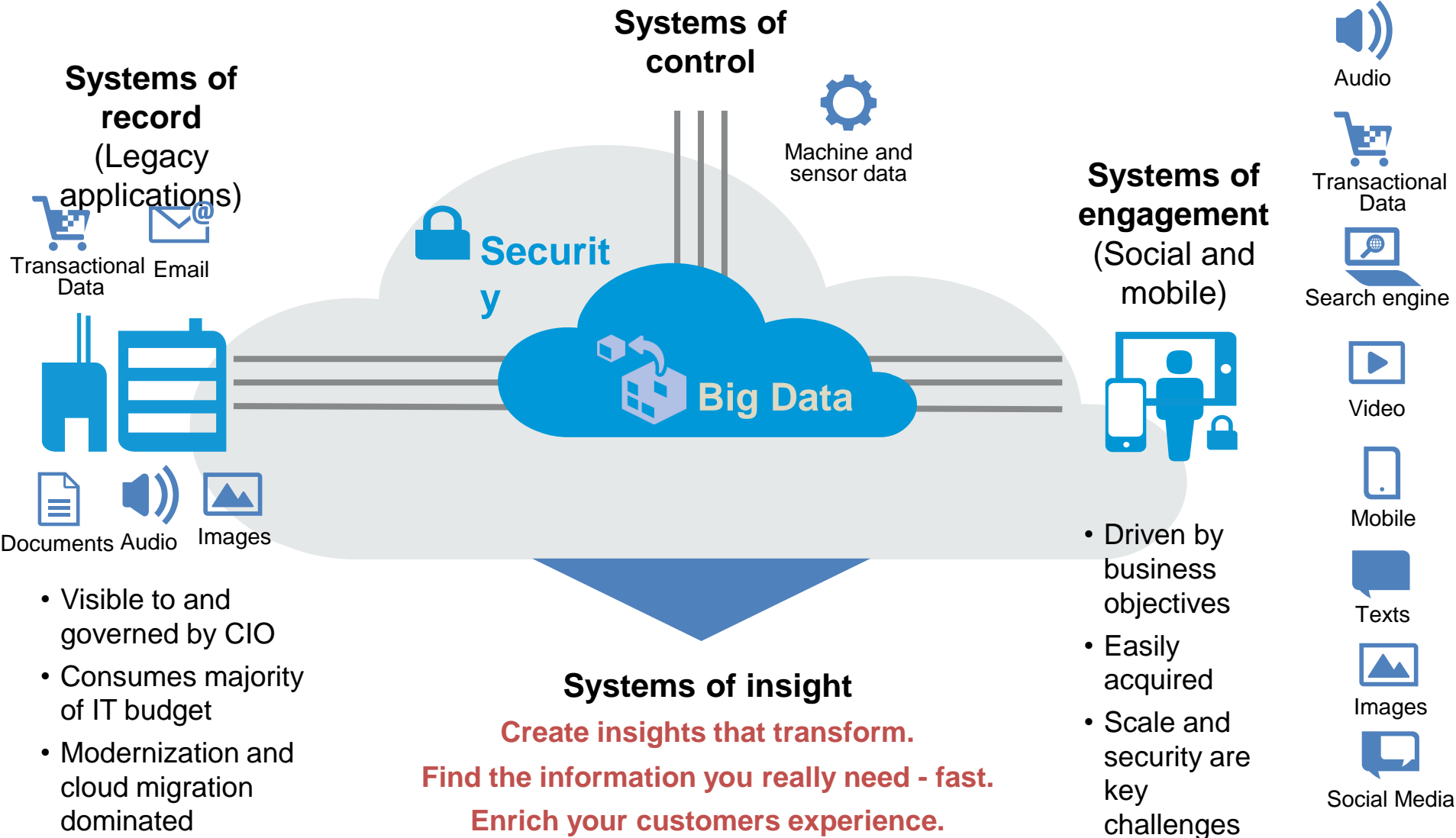EXTERNAL SOURCE

Hootsuite  TripIt  **Utilities**  iHandy  SolidFire  **Reference**  Scribd.  SmugMug
**Games**  Renren  Workday  Amazon  salesforce.com  **Finance**
Xactly  CloudSigma  **Navigation**  SuperCam  AppFog  Urban  **Travel**
Amazon  HP ePrint  CYworld  Twitter  Snapfish  **Productivity**  Facebook  Parse
Twitter  **Entertainment**  SuccessFactors  Zillabyte  Dragon Diction  GoGrid  LinkedIn  PingMe
Cookie Doodle  Yandex  RightScale  UPS Mobile  Bromium  Atlassian
myHomework  Tumblr.  Toggl  Jive Software  buzzd  **Lifestyle**
**News**  SmugMug  Atlassian  **Social Networking**  Amazon Web Services  Splunk
Fring  **Business**  Xing  Workbrain  Scanner Pro  LimeLight  box.net
Zynga  Khan Academy  MailChimp  Foursquare  Taleo  **Educa**
**Photo & Video**  MobileFrame.com  Answers.com  kaggle  Rackspace
**Sport**  **Music**

Largely unmanaged, ungoverned, unused...

APPLICATION

**KNOWLEDGE / INSIGHT** — BI (Modernization), Analytics, Knowledge Management, etc.

**INFORMATION** — Document Management, Content Management, Records Management, etc.

INFRASTRUCTURE

**DATA** — HDFS, Archiving, Backup & Recovery, Migration, Disaster Recovery, etc.

**STORAGE** — Server, Storage Arrays, Tiering, etc.

# The Changing Face of Governed Data

**Systems of record**
(Legacy applications)

Transactional Data    Email

Documents    Audio    Images

- Visible to and governed by CIO
- Consumes majority of IT budget
- Modernization and cloud migration dominated

**Systems of control**

Machine and sensor data

**Security**

**Big Data**

**Systems of insight**

**Create insights that transform.**

**Find the information you really need - fast.**

**Enrich your customers experience.**

**Systems of engagement**
(Social and mobile)

- Driven by business objectives
- Easily acquired
- Scale and security are key challenges

Audio

Transactional Data

Search engine

Video

Mobile

Texts

Images

Social Media

# A record is a **collection** of information

"As well as the content, the record should contain, or be persistently linked to,

or associated with, the metadata necessary to document a transaction, as follows:

...the structure of a record; the business context in which the record was created, received and used;  the links between documents, held separately but combining to make up a record

•Source: ISO 15489-1: 2001 Section 7.2.1

# Information Governance Definition

*"Information governance is defined as the processes, policies, standards, organization, and technologies required to manage and ensure the availability, accountability, quality, consistency, auditability, and security of information in an organization."*

Machine Data

Business Data

Human Information

Processes    Policy's    Standards    Organisation

Technology

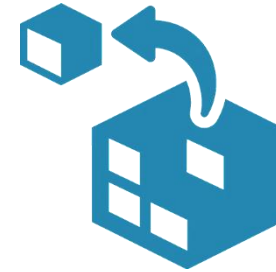Accessible    Context    Durable    Evidence    Quality    Reliable    Retained    Secure

# Why Data Governance for Big Data?



**Traditional**

**Big Data**

| Flexibility | Self-Service | • **Security** |
| | | • **Confidentiality** |
| | | • **Privacy** |
| | | • **Compliance** |

| Speed | Agility | • **Quality** |
| | | • **Trust** |
| | | • **Usability** |

# Examples

## Governance/Risk/Compliance of Sensor & Machine Data



- How is this data used?

- What business decisions are based on it?

- What were all the source data sets

- What elements need to be stored?

- For how long?

## Social Media and its impact on Information Governance



- How is this data used?

- What business decisions are based on it?

- What were all the source data sets

- What elements need to be stored?

- For how long?

## Governance, Ownership & Security of data in central hubs



- Who owns the data?

- How can it be used?

- Who has access?

- What are the compliance requirements around it?

- How do we maintain data lineage?

# Examples

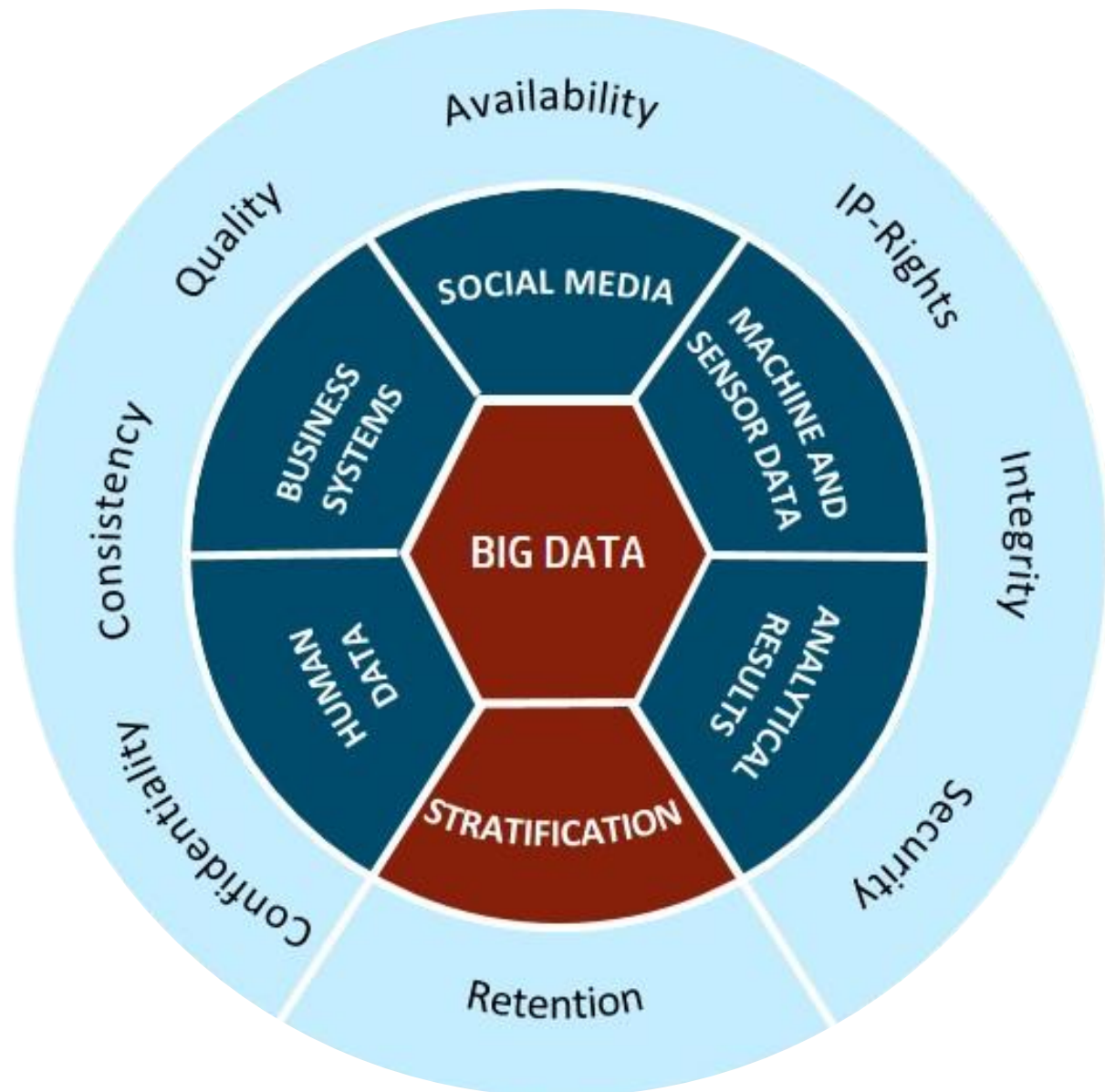## How do you Information Govern Organisational Big Data?



- Who owns the data?

- How can it be used?

- Who has access?

- How do we maintain data lineage?

-  What is the business risk if not managed?

# BREAKING IT DOWN – OUR FRAMEWORK

# GOVERNANCE IN THE WORLD OF BIG DATA

# THE 4 CORNERSTONES OF DATA GOVERNANCE

THE ORGANIZATION

STRATIFICATION

POLICIES & PROCESSES

ENABLING ARCHITECTURE

# Data Governance- The Organisation

**How you organise your governance, risk and compliance approach**
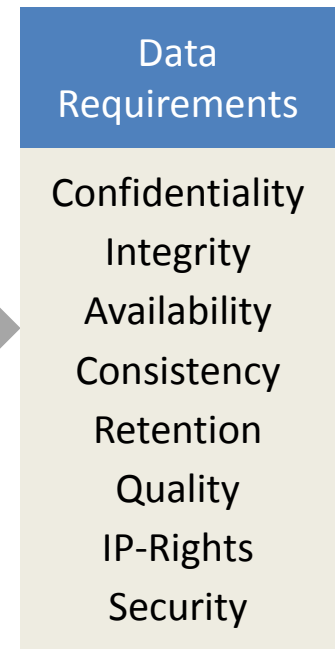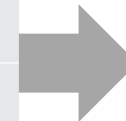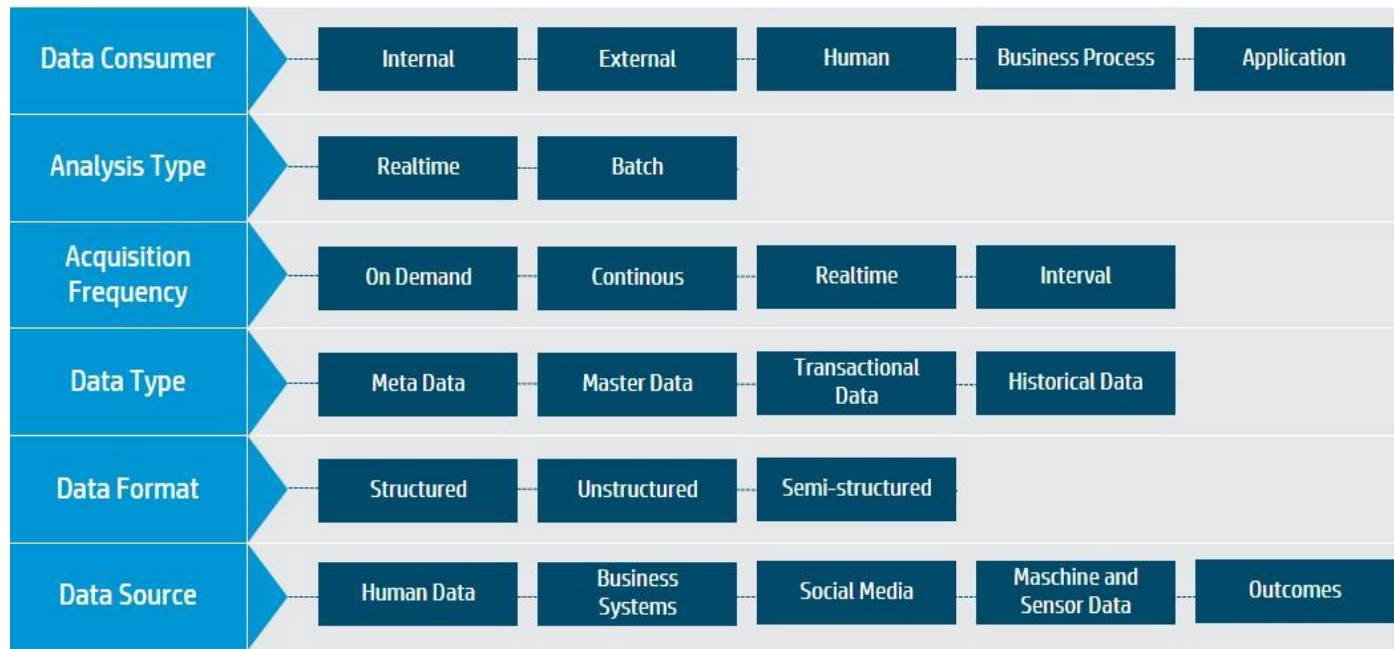


| Strategy & Plan | Pilot | Implement |
|:---:|:---:|:---:|

# Stratification Building Blocks

## How you stratify and classify your data

The Business Context can be categorized and aligned to the data sources. This will help to determine the appropriate classification method (atomic or composite).



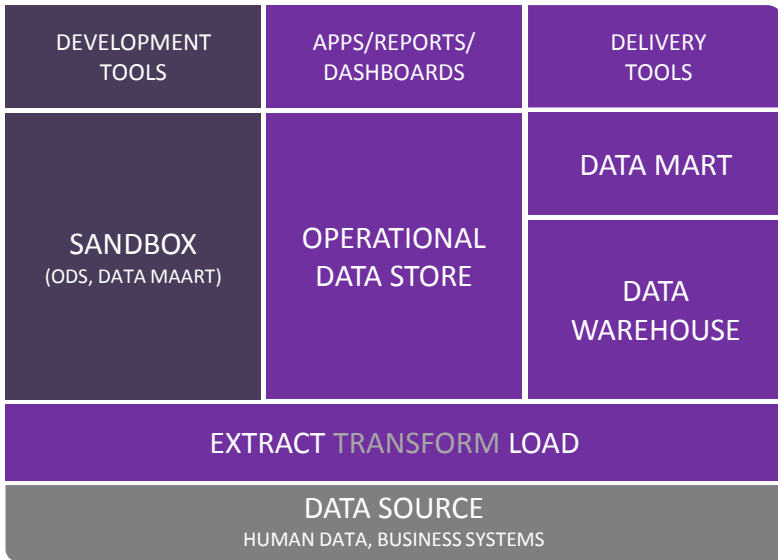| Data Consumer | Internal | External | Human | Business Process | Application |
| --- | --- | --- | --- | --- | --- |
| Analysis Type | Realtime | Batch | | | |
| Acquisition Frequency | On Demand | Continous | Realtime | Interval | |
| Data Type | Meta Data | Master Data | Transactional Data | Historical Data | |
| Data Format | Structured | Unstructured | Semi-structured | | |
| Data Source | Human Data | Business Systems | Social Media | Maschine and Sensor Data | Outcomes |

**Data Requirements**

Confidentiality
Integrity
Availability
Consistency
Retention
Quality
IP-Rights
Security

# Policies & Processes

## The policies & processes to be adopted and managed



DATA GOVERNANCE
FOR
BI MODERNIZATION

| Policy & Process Category | Confidentiality | Consistancy Quality | Availability IP-Rights | Integrity Security | Retention |
|---|---|---|---|---|---|
| Define & Approve | Process defines how the criteria and gradings for the policy will be identified, developed and approved | | | | |
| Implement, Enforce, Automate | How and with what Policies will be applied, including area of application | | | | |
| Monitor & Validate | Adherence to Policies and Processes, continous improvements | | | | |
| Communication & Education | Enablement and Training | | | | |

# Enabling Architecture - Hybrid DM

**The new world in which you will be governing new types of data**



Production, Business Critical Governance

| DEVELOPMENT TOOLS | APPS/REPORTS/ DASHBOARDS | DELIVERY TOOLS |
|---|---|---|
| | | DATA MART |
| SANDBOX (ODS, DATA MAART) | OPERATIONAL DATA STORE | DATA WAREHOUSE |

EXTRACT TRANSFORM LOAD

DATA SOURCE
HUMAN DATA, BUSINESS SYSTEMS



| VISUALIZATION TOOLS | APPS/REPORTS/ DASHBOARDS | DISCOVERY TOOLS |
|---|---|---|
| ANALYTICAL ENGINE | | |
| STREAMING ANALYSIS | DATA LAKE | |
| | EXTRACT LOAD TRANSFORM | |

DATA SOURCE
SOCIAL MEDIA, MACHINE AND SENSOR DATA

## TRADITIONAL BI

## HYBRID BI

# Questions?

## Gareth Martin

gareth.martin@hp.com
+44 7879 113 440
Big Data Analytics Lead - EMEA

## Grace Crossley

grace.crossley@hp.com
+44 7825 596 349
Information Governance Lead - EMEA

# Navigating Your GRC Journey

Thank You