1. Understanding Big Data

Definition & Characteristics

“Big Data is the collection of large and complex data sets that are difficult to process using “traditional” database management tools or data processing applications. Big Data is the new “raw material of business.”” – The Economist

BUT

As Dan Ariely, Professor at Duke University said: “Big Data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it.”
Understanding Big Data
Definition & Characteristics

Big Data generally refers to a set of technologies and initiatives involving data that is too fast changing (velocity), massive (volume) or too diverse (variety) for conventional technologies, skills and infrastructure to handle efficiently.

<table>
<thead>
<tr>
<th>Velocity</th>
<th>10% of the data in the world today has been created in the last 2 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume</td>
<td>Every day, we create more than 2.5 quintillion bytes of data</td>
</tr>
<tr>
<td>Variety</td>
<td>Diversity of data sources and formats</td>
</tr>
</tbody>
</table>

• Structured to semi-structured to unstructured

Analytics on almost all types of data
Virtualization architecture
Often distributed large-scale (cloud) infrastructures

Understanding Big Data
A different set up

• No “one to one” relationship of server to data storage
  • Reliance on virtualization architecture, needed to be able to draw from large content stores and archives as a single global resource

• Big Data environments often rely on distributed large-scale (cloud) infrastructures
  • A diversity of data sources and a high volume + frequency of data migration between different (cloud) environments

Understanding Big Data
Big Data Analytics

Collecting, organizing and analyzing
Large sets of data
To discover patterns and other useful information

Better decisions
Competitive advantage
Understanding Big Data

Analytics Opportunities

Big Data Analytics represents tremendous opportunities, both for the private and public sector:

• As a business asset: can serve to understand customers at a whole new level.
• To improve health care: can help deliver effective health care to patients faster and earlier (e.g. predictive medicine).
• To improve service: analytics can be used by governments to improve their citizens’ experience with administration by anticipating their needs.
• To strengthen security: Potential for new security insights and enhanced detection and prevention systems provided that the challenges and risks are properly mitigated.

Gathering information

Example:
Many apps request the user’s location in order to give more accurate search results (e.g. nearest restaurant or shop of a certain company).

Example:
The content of the emails sent and received through Gmail are scanned by an automated software of Google. This information is then combined with other information of the Google profile of the users to display more relevant ads.
Example:
Data gathered through online surveys are used by companies to gain more insight about their clients (e.g., average age of people consuming their products).

Example:
Several service providers combine the browsing history of an individual user with the knowledge gathered about other users following a similar browsing path to give more accurate search results or advertising (e.g., Amazon and FNAC's books' suggestions, YouTube's videos' suggestions, etc.).

Example:
Data about the purchases made by a customer with its loyalty card, sensors in a car to determine the driving style of an individual, etc.

Understanding Big Data
Gathering information

- Online surveys
- Browser history
- Other sources

Big Data in perspective: General Data Protection Regulation

- Privacy by Design
- Privacy by Default
- Privacy Impact Assessment
- Records of Processing Activities
- Data Security of Processing
- Breach Notification
- Data Protection Officer

2. Big Data vs. Security & Privacy
Creating entirely new challenges we have not encountered before:

- **Data linkages**: Powerful analytics solutions can link data sets to reveal someone's lifestyle, consumer habits, social networks and more – even if no single data set reveals this personal information.
- **Profiling**: The use of identifiable data to profile individuals in order to analyze, predict and influence their behavior.

---

**Big Data vs. Security & Privacy Security Challenges**

<table>
<thead>
<tr>
<th>BIG DATA</th>
<th>TRADITIONAL SECURITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access to huge volumes</td>
<td>Difficult to protect and monitor</td>
</tr>
<tr>
<td>Complex environments</td>
<td>Lack of granular audit trails</td>
</tr>
<tr>
<td>Variety of data</td>
<td>Difficulty to verify access</td>
</tr>
</tbody>
</table>
Big Data vs. Security & Privacy

Security Challenges

- **Big Opportunities**: Big Data is now often regarded as most critical enterprise asset, focusing attention on performance and collection of data, not security.
- **Big Attackers**: Big Data attracts a new class of hackers & attacks. Threat landscape has altered radically.

Opportunity example

Using Big Data to analyse, predict and prevent security incidents

Big Data provides the opportunity to consolidate and analyse logs automatically from multiple sources rather than in isolation

Potential for new insights and enhanced detection and prevention systems through continual adjustment and effectively learning “good” and “bad” behaviours

Challenge example

Securing the organisation and customers' information

Information classification and data ownership become more critical

Encryption and access controls based on data attributes rather than storage environment
3. Conclusive remarks

Conclusive remarks
From a classic data protection governance model to an agile one...

- The Big Data security challenges will require a more agile security governance model including:
  - More attention to detail: A holistic privacy/security strategy
  - A migration from point products to a more unified security architecture
  - Open and scalable Big Data security tools and approach
  - A strengthening of SOC’s (Security Operations Centre) data science skills
  - A more extensive leverage on external threat intelligence
  - A more pragmatic focus on (breach) incident as well as identity and access management

Conclusive remarks
From a classic data protection governance model to an agile one...

- In addition, a more agile governance model should address the main privacy challenges of Big Data:
  - Invest in IT security governance, not only security products
  - Manage the security/privacy paradox and use an integrated security/privacy approach (e.g. monitoring versus anonymization)
  - Clearly define privacy (and security) responsibilities
  - Ongoing monitoring and audits (e.g. privacy impact assessments)
  - Focus on the obtaining consent of the data subject: opt-in, not opt-out
  - Make sure that processing remains compatible with purpose of collection (e.g. "secondary use" issue)
  - Engage in harmonization and standardization
  - Transparency towards data subject regarding its (GDPR) rights
Conclusive remarks
A holistic risk based approach across the (big) data lifecycle

- Effective data protection tools across the data lifecycle to allow an enterprise to tailor policy in a way that keeps information safe, yet available to those authorized to access it.
- Without knowing the lifecycle of data flowing through your organisation, it is impossible to be sure that it is all managed appropriately.
- Creating a personal data inventory and/or personal data flow maps will allow to understand and analyse the scope of privacy in your organisation.

General risks

- Effective data protection looks across the data lifecycle to allow an enterprise to tailor policy in a way that keeps information safe, yet available to those authorized to access it.
- Without knowing the lifecycle of data flowing through your organisation, it is impossible to be sure that it is all managed appropriately.
- Creating a personal data inventory and/or personal data flow maps will allow to understand and analyse the scope of privacy in your organisation.

Thank you for your attention Any questions?