Privacy & Big Data:

5 Keys to Marrying Up Big Data and Privacy by Design

IAPP Europe Conference
Ronald Koorn & Joris Lindenhovius
12 December 2013
Agenda

1. What is 'Big Data'?
2. Privacy Implications
3. Privacy by Design
4. KPN Big Data Case
What is ‘Big Data’?

Ronald Koom
‘Big Data’ context

- Big Data & Privacy
- Individual vs. Business
- Compliance
- Security
- Data Quality
- Business Process Management
- Corporate Social Responsibility
- Legal
- Integrity & Fraud Protection
- Controls/Auditing
- Records Mgmt. / Data retention

- Individual vs. Business
- Compliance
- Security
- Data Quality
- Business Process Management
- Corporate Social Responsibility
- Legal
- Integrity & Fraud Protection
- Controls/Auditing
- Records Mgmt. / Data retention

© 2013 KPMG Advisory N.V., registered with the trade register in the Netherlands under number 33263682, is a subsidiary of KPMG Europe LLP and a member firm of the KPMG network of independent member firms affiliated with KPMG International Cooperative (“KPMG International”), a Swiss entity. All rights reserved.
What is Big Data?

Evolution or revolution?
What is ‘Big Data’?

Big data is high-volume, high-velocity & high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight & decision-making (Gartner)

...every day more than 2.5 exabytes ($2.5 \times 10^{18}$) of data is created
What is ‘Big Data’?

The amount of data is growing **40 times** as fast as the world population.

**Traditional:** Use customer data to run bricks-and-mortar operations. Analyse website metrics to understand e-commerce operations.

**Rise of big data with smart phones & social media:** Need for more agile, relevant customer engagement to differentiate from the crowd.

Diverse and scattered data across silos contain **VALUE**

Traditional data approaches are not **UNLOCKING** the trapped value

**Areas for investment in Big Data projects**

- Customer Service
- Marketing
- Innovation
- Risk and Profit Mgt
Examples of Big Data applications: Analysis of Earnings per Antenna

Measurement of revenue per antenna cell at one single day. Green dots indicate the antennae with relative large amount of revenue, while the red dots indicate antennae with relatively low revenue. Dots in the sea indicate antennae on oil rigs.

Measurement of the average revenue per antenna versus the amount of connections. The size of the dots indicate the total sum of revenue. Outliers can be spotted easily to find out issues with costly antennae. Based on this information it is possible to prioritize maintenance of infrastructure by identifying the most important antennae.
Examples of Big Data applications: Royal Coronation & Crowd Control

Widget (left) from *waarisdekonning.nl* that monitors crowd density during the coronation of the new King in Amsterdam on the 30th of April 2013. The crowd density was determined by projecting the mobile traffic load on antennae of a Telco on any given location in the city. The website (below) also showed tweets, including sentiment, on the map.

Animation of tweets in The Netherlands regarding Project X on 21st of September 2012: a party in Haren (North-East of Holland) organized through an accidental open invite on Facebook resulted in riots.
Examples of Big Data applications: Modeling Mortgage Risks

Risk modeling for a bank's collection of mortgages based on life-changing events. By combining public information on an online house-selling website (funda.nl) with internal bank data on transactions, a model has been built to describe and predict the length of mortgages at that bank (blue curve).

Analysis of the difference in house-selling price and the mortgage at the bank. The red curve corresponds to customers in so-called ‘special treatment’ for negative financial reasons, while the blue curve corresponds to normal customers that have not been in this special treatment.
Examples of Big Data applications: Sentiment Analyses

Sentiment analysis of tweets (with GPS information) that include the brand name as hashtag (i.e. #<brand>). It can be used to measure for example the customer experience of services offered by the brand, or the response of people to TV commercials related to the brand in a specific region and over time.

Sentiment analysis of tweets that include the brand name as hashtag (i.e. #<brand>) over time. Note that the peak on 19th of August coincides with a nation-wide service disruption that was compensated by offering a 3-days free service offering.
Agenda

- What is 'Big Data'?
- Privacy Implications
- Privacy by Design
- KPN Big Data Case
“Big data is centered inherently around the individual. The dependency has the potential to create significant negative consequences.”

Privacy & Big Data – ISACA whitepaper, August 2013

Understanding customer interactions

Amazon £27.95
All-Bar-One £112.95
@Me Oh boy I’m tapped
ATM £100.00 cash
Experian £1000 Credit Check
Hey, they approved my credit increase
Refer to bank no funds
McDonald’s £11.50
eBay £250.00
Privacy Implications (cont.)

Objectives

To understand public opinion regarding online privacy in nine countries globally.

- **Europe**: UK, Germany, France, Spain.
- **Non-Europe**: India, Japan, South Korea, Brazil, Australia.
- **10,354 interviews in total**

Consumers are concerned about (online) Privacy

Consumers in India, followed by Brazil and Spain are the most likely to say they are concerned.

Germany is the only country where a majority say they are unconcerned.

Mixed views exist on big companies collecting personal data

Consumers in South Korea, UK, Australia and France are the most critical.

Consumers in Brazil, India and Spain are the most sympathetic.

http://www.bigbrotherwatch.org.uk/home/2013/06/new-research-global-attitudes-to-privacy-online.html#more-5479
## Privacy Requirements for Big Data projects

<table>
<thead>
<tr>
<th>Privacy Principles</th>
<th>Big Data Implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPA notification</td>
<td>DPA to be notified (depending on use)</td>
</tr>
<tr>
<td>Transparency</td>
<td>Data subjects (and work council) to be notified (esp. autom. decision-making)</td>
</tr>
<tr>
<td>Purpose Binding</td>
<td>Limit further incompatible use</td>
</tr>
<tr>
<td>Legitimate Grounds</td>
<td>Consent to individual profiling</td>
</tr>
<tr>
<td>Data Quality</td>
<td>Match &amp; enrich in/external data sources</td>
</tr>
<tr>
<td>Rights of Data Subjects</td>
<td>Data subjects to be notified and right to object / opt-out</td>
</tr>
<tr>
<td>Security</td>
<td>Datawarehouse is single point of failure/success; strong access controls. Logging, monitoring &amp; data destruction required</td>
</tr>
<tr>
<td>Outsourcing / Onward transfer</td>
<td>Data ownership, contracts</td>
</tr>
</tbody>
</table>
Agenda

- What is 'Big Data'?
- Privacy Implications
- Privacy by Design
- KPN Big Data Case
Privacy by Design: Privacy Impact Assessment

Phase 1: Determine Scope & Approach

Phase 2: Inventory Dataflows, Sources, Elements & Results

Phase 3: Identify Applicable Legislation & Expectations

Phase 4: Map Regulatory Requirements on Big Data Analytics

Phase 5: Identify Big Data Privacy Risks

Phase 6: Select Privacy by Design solutions

Phase 7: Implementation & Auditing
Privacy by Design: Security is insufficient for Big Data Privacy

Privacy
- Privacy Policy
- Notification
- Subject Rights
- Purpose Binding
- Legitimate Grounds
- Data Quality
- Lawful transfer (incl. outside EU)

Security
- Security Policy
- Classification
- Logical Security
- Physical Security
- Availability
- Compliance
- Incident Mgmt.

Security Organisation
- Personnel Security
- IT Services Mgmt.
- System Development
Privacy by Design: Solutions & Effectiveness

- General PET controls
- Data separation
- Privacy management systems
- Anonymisation

**PRIVACY BY DESIGN !**
Privacy by Design: Anonymization vs. Pseudonymization

**Anonymization:**

- No privacy legislation applicable
- Totally anonymous is challenging with substantive data sets
- Less useful for combining internal and external data
- Limited security requirements remaining after de-identification

**Pseudonymization:**

- EU Regulation still applicable, in some countries the EU Directive is not
- Useful for longitudinal data analytics & research
- Close to the data source
- Use of Trusted Third Party effective solution
- No indirect inheritance possible (non-reversible, i.e. one-way hashing)
- Susceptible for inference attaches (potential re-identification)
Effects of upcoming EU Data Protection Regulation on Big Data

- US-based data warehouse with EU citizens within scope
- Privacy Impact Assessment mandatory
- Data minimization & profiling only achievable with anonymization / pseudonymization
- Separate consent for data analysis may be required (incl. parental consent for minors)
- Extensive data breach notification required
- Data portability easier to provide
- Right to be erased (for all registrations, incl. transferred data)
- Privacy by Design required
- Penalties can be prohibitive
Discussion Questions

1. What do you consider to be key (personal) data assets of a telco?

2. How can a company leverage Big Data opportunities for customer satisfaction / experience?

3. Do you see the Big Data opportunity a natural area for EU-based telcos to divulge in or leave it to US-based (multi-channel) players (Facebook, Apple, Google, etc.)?

4. What are challenges for traditional telco in upcoming data market (information-driven society)?

5. Which solutions do you see to overcome the Privacy hurdles of Big Data?

6. How to enforce and show Privacy Compliance on a continuous basis?
Take Aways

Big Data projects are better safeguarded by following 5 keys:

1. Accountability / data Governance
2. Privacy Impact Assessment
3. Anonymization / Pseudonymization
4. Organizational, contractual, technical controls
5. Communication Strategy & Consent
Further discussion input?

Big Data is too important to leave to (data) nerds

Privacy is too important to leave (solely) to lawyers
Agenda

- What is 'Big Data'?
- Privacy Implications
- Privacy by Design
- KPN Big Data Case
Thank you

koorn.ronald@kpmg.nl
joris.lindenhovius@kpn.com

KPMG IT Advisory
P.O. Box 43004
3540 AA Utrecht
The Netherlands
Tel. +31 (0)30 658 2150