Big Issues with Big Data

Jeff N. Jensen

Office of Legal Affairs

September 29, 2016

Big Data is Everywhere

- Entertainment industry uses big data to create programming to meet customer interests

- Brokers gather employment data and sell it to financial institutions, debt collectors, etc.

- Retailers identify repeat customers and target them with product marketing

Big Data

- The potential uses and benefits of Big Data appear <u>unlimited</u>

- But are there risks …
  - For companies and organizations aiming to unlock the potential of Big Data?
  - For individuals whose information is collected, analyzed, disclosed and acted upon?

What is Big Data?

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.

Dimensions of Big Data

Volume

Velocity

Variety

(No dimension is static, all continue to change)

Volume

The sheer amount of data generated or data intensity that musts be ingested, analyzed, and managed to make informed decisions.

Contributing factors:

- Recent <u>explosion</u> of data creation

- Data storage costs are <u>decreasing</u>

Velocity

How fast data is being produced and changed and the speed with which data must be received, understood, and processed.

Velocity

The velocity of data creation today is <u>astonishing</u>!

- From the beginning of recorded time until 2003, we created 5 billion gigabytes (5 exabytes) of data

- In 2011, the same amount was created every 2 days

- Today, the same amount of data is created in less than 10 minutes

Variety

- Data analysis traditionally required a structured data set

- Big Data technologies allow analytics to be run across databases of different structures or even unstructured data consisting of text, sensor data, audio, video, graphic, etc.

- New insights are found when analyzing these data types together

What does Big Data mean to companies or other organizations?

1. Organizations can access unimaginable amounts of structured and unstructured data, gathered both internally and through external resources

2. Business goal to harness and understand the relationships within and between data to gather insights and improve their processes

What does Big Data mean to companies or other organizations?

3. Need to employ specialized tools and data scientists to capture and analyze in a way that is valuable to the organization

4. Must also analyze and address the potential limitations, risks and legal issues associated with the collection, analysis and use of Big Data

Who uses Big Data?

- Banking & Finance
- Education
- Government
- Weather
- Health Care
- Manufacturing
- Retail

Legal Considerations & Challenges

Consumer Reporting Agencies (CRA)

- Compile and sell consumer data reports

- Used to determine consumer eligibility for certain benefits and transactions

- FCRA applies to Consumer Reporting Agencies

Fair Credit Reporting Act (FCRA)

Must implement reasonable procedures to ensure maximum possible accuracy of consumer reports

- Provide consumer <u>notice</u> of their report
- Provide consumer with <u>access</u> to their information
- Provide consumer an <u>opportunity to correct errors</u>

What about Data Brokers?

- Compile information from social media and other nontraditional sources

- May or may not be considered a "consumer reporting agency" subject to the FCRA

Legal Considerations & Challenges

Spokeo – an online data broker

- assembled personal information from hundreds of online and offline data sources, including social networks, and merged the data to create detailed personal profiles (including name, age range, hobbies, ethnicity, and religion)

Legal Considerations & Challenges

Spokeo – an online data broker

- marketed their data profiles for use by human resources departments to aid them in making hiring decisions

- web disclaimer said "we are not a CRA" so don't use our data for eligibility purposes

FTC disagreed

Legal Considerations & Challenges

Instant Checkmate

- Data broker that assembled consumer data for employment and tenant screen purposes
- Advertised its services on its website and through blog posts
- Also used Google AdWorks campaign to display ads for its services with search results when consumers sought background checks on nannies, babysitters, maids and housekeeper
- web disclaimer said "we are not a CRA" so don't use our data for eligibility purposes

Legal Considerations & Challenges

FTC disagreed

- Ruled that Instant Checkmate had advertised it background check services

- Found Instant Checkmate was subject to, but had failed to comply with, the FCRA

- Determined web disclaimer to be insufficient and ineffective

- Consent decree ordered future compliance and payment of $550,000 in civil penalties

Legal Considerations & Challenges

Users of Consumer Reports also have FCRA Obligations

- Must provide consumer with "adverse action" notice if the consumer report information is used to deny them a covered benefit (credit, employment, insurance, housing, etc.)
- Must provide "risk-based pricing" notice if they charge consumer more to obtain a covered benefit based on consumer report information

The purpose of these notices is to inform and enable consumers to check their consumer reports and correct any inaccuracies

Legal Considerations & Challenges

Time Warner Cable

- Used consumer report information to determine whether to require a deposit on a consumer's cable bill
- FTC determined that consumers who were charged a deposit should have received a "risk-based pricing" notice informing them that the charge was based on information in their consumer report
- Consent Order included $1.9 million in civil penalties and required future compliance with Risk-Based Pricing Rule and

Legal Considerations & Challenges

Sprint

- Failed to provide "risk-based pricing" notices to consumers who were placed in a program for customers with lower credit scores and charged an extra monthly fee

- Consent order included a $2.95 million penalty and required Sprint to provide timely notice to consumers placed in the program

Legal Considerations & Challenges

Question: If a retailer relies upon a general analytics report from a data broker, obtained through a web search, to deny a consumer credit, can the consumer request disclosure of the nature of the information leading to the denial?

Lesson: Retailers and consumers should be mindful of the law concerning the use of big data analytics to make consumer eligibility determinations.

Big Data and Privacy

A significant legal challenge with Big Data is privacy!

- When it comes to consumer marketing, the potential for Big Data in enormous.

- Given the enormity, how best can privacy protections be observed?

Big Data and Privacy

Privacy protections include the concepts of:

- Notice and Awareness, and

- Choice and Consent

Big Data and Privacy

Can goals of Notice/Consent be circumvented?

- The Big Data Ecosystem is Complex

- Practical limitation related to the use of written privacy polices

Big Data and Privacy

- Problem 1: The consumer may not understand where her or his personal information may end up, and that it could be combined with other existing profile data in a manner that reveals more about the person than contemplated at the time of disclosure.

Big Data and Privacy

- Problem 2: The data subject lacks an understanding of the interpretations, inferences and deductions that may be drawn from her or his combined data using Big Data mining techniques and analytics.

In this Big Data context, many would argue that data subjects have limited awareness and ability to provide meaningful, informed consent.

Big Data and Privacy

Does Big Data allow for "access/participation"?

- This principle concerns a data subject's ability to access her/his personal data in order to ascertain whether it is accurate and complete.

- Necessary to allow an individual the opportunity to correct inaccurate information.

Big Data and Privacy

Do Not Track, Do Not Target, and Do Not Collect?

- Another privacy-related area impacted by Big Data

- For the advertising industry, "Do Not Track" refers to the use of consumer data for purposes of targeted advertising.

- In contrast, privacy advocates and the FTC believe that DNT prohibitions should encompass not only the targeting of individuals, but also the collection of personal information from individuals.

De-Identification or Anonymization

- One technique for mitigating privacy-related risks associated with Big Data is de-identification or anonymization.

- Data sets that are de-identified have had key information stripped away, in order to prevent others from individually identifying the persons to whom the data set relates.

- The technique allows organizations to work with Big Data sets while mitigating privacy concerns, and has been used in healthcare, banking, finance and on-line advertising.

De-Identification or Anonymization

Many regulatory regimes, including finance and health care, recognize the concept of de-identified personal information

- Under Gramm-Leach-Bliley regulations, "personally identifiable financial information" does not include information that does not identify a consumer, "such as aggregate information or blind data that does not contain personal identifiers such as account number, names, or addresses."
- HIPAA offers two methods for health data: expert determination and "safe harbor" de-identification which involves removing eighteen types of identifiers.
- European data protection laws: "anonymization of data should exclude any possibility of individuals to be identified, even by combining anonymized information."

De-Identification is not Full Proof

If de-identification is not performed properly, it may be possible to re-identify individuals in an anonymized data set.

- Netflix Contest

- Massachusetts Healthcare Data

How to Mitigate Risk?

Organizations desiring to de-identify and anonymize their data sets should consider several questions to help understand and mitigate potential privacy and organizational risks, including:

1. What are the purposes, risks and benefits of de-identifying and using or disclosing the data, and do the benefits outweigh the risks?

How to Mitigate Risk?

2. Is the data truly anonymized? What information will be exposed if the data is re-identified? Is it worth investing effort to verify anonymization?

3. What is the risk to the organization if the data is re-identified? A costly data breach notification, or a lawsuit or regulatory action?

Conclusion

It is increasingly common for companies to collect, data mine and analyze large data sets in order to further their business or organizational interests. Big Data analytics is already the norm for many organizations, and the trend will only continue in the coming years as more and more data is produced and stored, and as stronger and more predictive tools and processes are developed to understand that data.

Proactively dealing with privacy and other legal issues can help organizations more safely leverage Big Data while still retaining customers (or clients), and avoiding reputational harm, litigation and regulatory scrutiny.

Michael Korvink

Senior Data Scientist

Premier, Inc.

Big Data Analytics and

Health Care Organizations