Thinking small about big data:
Privacy considerations for the public sector

Shaun Brown

Partner, nNovation LLP

March 30, 2016

nNovation LLP

Thinking small about big data: objectives

- Consider big data as a concept
- Focus on predictive analytics
- Emphasize practical over policy
- Understand the legal landscape (*Privacy Act*)
- Key takeaways:
  - Three step analytical framework
  - Practical considerations when using predictive analytics

What is big data?

- The “three Vs”:
  - High-Volume
  - High-Velocity
  - High-Variety
- About linking datasets and identifying patterns
- Fueled by ICTs, cheap storage, applications that create “digital exhaust”

What is big data?

- A problem to be solved:
  - *"…a term for data sets that are so large or complex that traditional data processing applications are inadequate. Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, querying and information privacy"* (Wikipedia)

What is big data?

- (Business) Opportunity:
  - *"…information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation*." (Gartner)

What is big data?

- Our worst nightmare?
  - Systematic discrimination
  - Loss of autonomy
  - Rise of the machines
- Difficult policy issues about regulation of technology

Predictive Analytics

- Predictive Analytics: thinking *a bit* smaller
  - “*statistical techniques from predictive modeling, machine learning, and data mining that analyze current and historical facts to make predictions about future or otherwise unknown events*” (Wikipedia)

Predictive Analytics

- It is already everywhere:
  - Credit scoring
  - Online behavioural advertising (OBA)
  - Recommendation engines (e.g., Amazon, Netflix)
  - Human resources: employability and retention
  - Crime prevention and anti-terrorism: predict when and where crime is likely to happen.

Legal landscape: *Privacy Act*

- Applies to the collection, use and disclosure of Personal Information (PI) by federal government institutions (departments, ministries, parent Crown Corporations)
- PI very broadly defined:
  - Information about an identifiable individual in any form
  - "*Information will be about an identifiable individual where there is a serious possibility that an individual could be identified through the use of that information, alone or in combination with other available information*." (*Gordon v. Canada*, Federal Court)
  - Includes things such as internet protocol address

Legal landscape: *Privacy Act*

- Authority: Can only collect PI if it relates directly to an operating program or activity
  - Contrast to private sector which relies on consent & reasonableness
- Identify Purposes: individuals must be informed of purpose for collection (some exceptions)

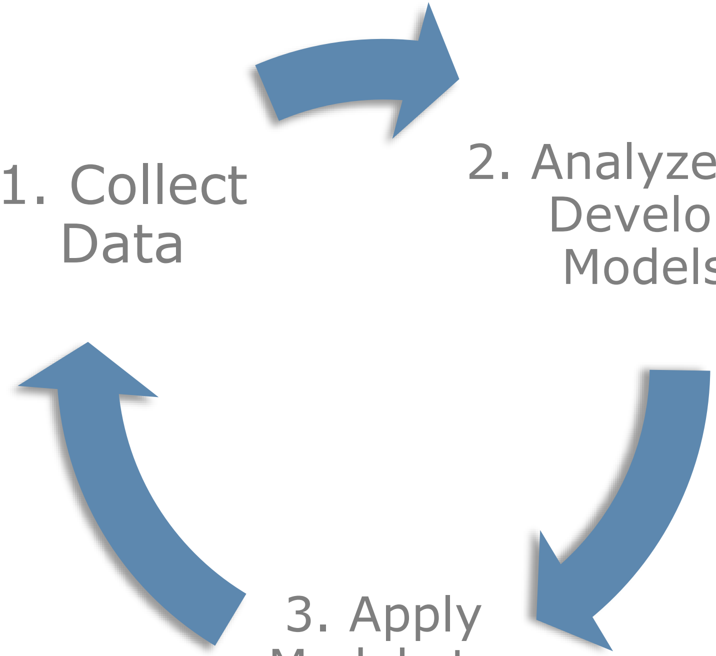
Legal landscape: *Privacy Act*

- Additional requirements where PI used for an "Administrative Purpose": *Decision making process that directly affects an individual*
- Direct Collection: PI collected for administrative purpose must be collected directly from individual (some exceptions)
- Retention: PI used for administrative purpose must be retained long enough for individual to access (minimum 2 years)
- Accuracy: Must take reasonable steps to ensure PI used for administrative purpose is as accurate and up-to-date as possible

Legal landscape: *Privacy Act*

- PI can only be used for the purpose for which it was collected or a consistent purpose (some exceptions)
- PI can only be disclosed with consent (some exceptions)
- Access Requirements:
  - PI and non-PI is subject to access under *Privacy Act* and *Access to Information Act*, respectively

Understanding PA: three step framework



1. Collect Data

2. Analyze and Develop Models

3. Apply Models to Individuals

- Highly simplified understanding of most common applications
- Applies to public sector orgs. looking to apply predictive analytics to a database of individuals
- Case study: *Beware* Threat Rating System used by Law Enforcement Agencies (LEAs)

Step 1: Data Collection

- What is feeding the “machine”? Remember the three “Vs” (volume, velocity and variety)
- Are predictive models based on data collected in the past (or from other sources)? e.g., credit scoring systems
- Is *your* data feeding the machine? If so:
  - Who has access to data?
  - Is it being used by the vendor?
  - Other customers of the vendor?
- Remember restrictions on disclosure of PI
- Data should be rendered so it is not PI (e.g., de-identified) whenever possible

Step 1: Data Collection

- Beware
  - Crawls billions of records in commercial and public databases
  - Social media and other “internet chatter”
  - Court documents (for records of criminal convictions)?
  - Data provided by LEAs?

Step 2: Analyze Data and Develop Models

- Have models already been created?
- Is your PI being used to develop models?
- If so:
  - Are the research questions and goals clearly defined? Open-ended exploratory research involving PI should be avoided
  - Is research for or consistent with purpose for which PI was originally collected?

Step 2: Analyze Data and Develop Models

- Beware:
  - Threat rating system uses proprietary algorithm to predict threat based on available information
  - Threat rating formula not publicly disclosed
  - Presumably cross-references incidents of criminal behaviour with various other information about an individual
  - Feedback from LEAs would be helpful in validating and improving models

Step 3: Apply Predictive Models

- Individuals are scored, rated, categorized and treated accordingly
- Involves collection and use of PI about an individual
  - Need PI to feed the model
  - The score is new PI that is created

Step 3: Apply Predictive Models

- Beware:
  - LEA enters name, address and/or cell no. of individual into app (on computer or mobile device)
  - Beware analyzes all available data about that individual
  - Applies model to give a “threat rating”
    - Do ratings already exist? Created in real-time?

Step 3: Impact on individuals

- Does it result in or contribute to an administrative decision?
  - If so, how can we verify accuracy?
  - Is information retained long enough (2 years)?
- Does the outcome become part of a permanent record? Who might see this in the future?
- Does it discriminate, either directly or indirectly?
- Is the final decision made by a human (e.g., model is merely for filtering purposes)?

Summary and Practical Considerations

- The *Privacy Act* imposes few restrictions on which applications can be used
- Predictive Analytics require information about real people: is your data being used?
- De-identification is crucial
- Predictive Analytics is much more than a data collection issue: can result in uses, disclosures, and administrative decisions

Summary and Practical Considerations

- Predictive Analytics can be complex: you need to understand what is happening and the privacy impacts
- Do not be afraid to push back on vendors for more information
- How will the public be informed?
- Privacy Impact Assessments are essential:
  - Identify and mitigate privacy-related risks
  - Documenting business processes data flows
- About more than just privacy:
  - Must consider human rights issues (s. 15 of the *Charter* right to equal treatment without discrimination)
  - Sometimes it can be just plain creepy (Target)

Resources

- Tene, Polonetsky, *Judged by the Tin Man: Individual Rights in the Age of Big Data*, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2311040
- *TBS Standard on Privacy and Web Analytics*: http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=26761
- *Privacy Commissioner Policy Position on OBA*: https://www.priv.gc.ca/information/guide/2012/bg_ba_1206_e.asp
- Cavoukian & El Emam, *De-identification Protocols: Essential for Protecting Privacy*, Jun. 25, 2014: https://www.ipc.on.ca/images/Resources/pbd-de-identifcation_essential.pdf
- Beware: https://www.west.com/safety-services/public-safety/powerdata/beware/#

About nNovation LLP

- Ottawa-based, boutique firm that specializes in regulatory matters, including privacy, access to information, advertising, and competition law
- Our clients include many government institutions at the federal and provincial levels, in addition to private sector companies, not-for-profit entities and industry associations
- We regularly provide a variety of services to public sector clients, either as consultants or legal counsel, including:
  - Privacy Impact Assessments (over 150 completed)
  - Privacy Management Frameworks
  - Legal opinions
  - Draft outsourcing agreements
  - Advise on and assist with specific privacy and access to information-related matters on an as-needed basis

Shaun Brown

sbrown@nnovation.com

613-656-1297