BIG DATA PROJECT VETTING TO ASSURE FAIR AND INNOVATIVE DATA USE

April 2016
KEY WORKSHOP MATERIAL

You can find a blog post with links to this key material related to the workshop on the IAF website at http://informationaccountability.org/gps-iaf-workshop/. The material includes:

• Workshop Agenda
• Workshop Hypothetical
• Slide Presentation by Marty Abrams and Lynn Goldstein
• Slide Presentation by Joyce Yeager
• IAF Big Data Ethics Initiative Overview
• A Unified Ethical Frame for Big Data Analysis (Part A)
• Enforcing Big Data Assessment Processes (Part C)
• Contextual Assessment Worksheet for Marketing (Part D)
• OPC 2/16 Internet of Things Research Paper
• Big Data and Analytics: Seeking Foundations for Effective Privacy Guidance
WORKSHOP AGENDA

9:00 AM  Basics on a 360° Review – Marty Abrams, IAF
9:30 AM  Key Values for Assessment – Lynn Goldstein, IAF
9:50 AM  Enforcement Agency Oversight in 2018 – Joyce Yeager, Office of the AG, Missouri
10:10 AM  Enforcement Outside the United States – Marty Abrams
10:15 AM  Translating Values into Processing Steps – Sheila Colclasure, Acxiom, and Lynn Goldstein
10:30 AM  Break
10:50 AM  Introduction of the Part D for Digital Marketing – Lynn and Sheila
11:00 AM  Small Group Exercise
  • Group A – Bryon Donovan, Acxiom, Joyce and Marty
  • Group B – Phil Ziperman, Office of the AG, DC, Sheila, and Lynn
12:10 PM  Group Report Outs
12:40 PM  How Does Fairness Touch on our Exercise – Joyce
12:50 AM  Closing Q & A
FOUR REASONS TO DO ETHICAL INFORMATION IMPACT ASSESSMENTS (“IIAS”)

1. IIAs provide the confidence that allows my organization to use big data

2. IIAs help us understand the data quality issues related to big data use

3. In a world with less legal certainty, more data breaches, and greater penalties, my organization has IIAs documentation to defend our use of big data when questions arise
   a) IIAs prepares my organization for the external big data oversight that will likely come in the United States
   b) IIAs, as part of a more mature, organizational information governance system, create a pathway to compliant big data in markets such as Europe, Latin America, and Asia

4. As privacy increasingly becomes a competitive differentiator, IIAs help us maintain customer loyalty and develop consumer trust
WHY WE ARE HERE

• When big data hits your privacy impact assessment process, you will need a version update
  • It will assess the ethical dimensions of your use of advanced analytics, not just the mechanics of privacy compliance
  • It will form the basis for legitimate data use outside the United States
    • EU legitimate interests
    • Canadian implied consent
  • It will be the foundation for Accountability 2.0
YOUR FACULTY

- Marty Abrams, IAF
- Sheila Colclasure, Acxiom
- Bryan Donovan, Acxiom
- Lynn Goldstein, IAF
- Joyce Yeager, Office of the Attorney General, Missouri
- Phil Ziperman, Office of the Attorney General, District of Colombia
WE WILL DEAL WITH CULTURAL DIFFERENCE

• European law differentiates privacy and data protection
  • Privacy protects individual autonomy and family life
  • Data protection assures fair processing even if individual control is not fully effective

• Many legal systems integrate the two into information privacy

• Generally, most legal systems cover the processing of personal data, while the U.S. focuses more on the application
  • Generally, thinking with data is not governed by law in the U.S.
CORRECTING FOR GLOBAL DIFFERENCES

• Big data processes may be divided into discovery and application phases
  • Discovery is very similar to research – knowledge creation
  • Application is making use of the new insights

• In the U.S., using data to discover new insights has very few limitations

• In most every other regime, permission is required for the processing and compatibility of the data for the processing

• IAF assessments are designed to correct for those differences
GETTING INTO IT

Think about your last doctor visit:

- She had a PC, your records were digital, and all scans and x-rays were visible
- She could cascade through the data and visuals and match to your issues
  - She now orders tests based on probabilities
- Tomorrow, machine learning will prompt additional questions
  - Here come the genomes
- The day after, your health monitor will send real-time data to systems that will trigger intervention remotely
- Benefits? Yes
- Risks? You bet
- To be duplicated in every sector, industry, social realm
ASSESSMENT PROCESSES ARE BASED ON LEGACY CONCERNS

• Yesterday
  • Early thinkers thought the best way to protect individuals was by assuring individual autonomy
  • Led to U.S. Privacy Act and OECD Guidelines
  • Made sense based on pre-database mainframe technology
    • Data provided mostly by the individual
    • Systems and data were the same

• Today
  • Generation after generation of new technological shocks
    • Today, it is observational world – think Internet-of-Things
  • Advanced analytics drive insights which become inferred data
  • Autonomy – individual consent – alone no longer protects against harms
  • Ken Cukier: “more is not just more it is different”
  • Yet, privacy assessments are based on 1970’s work
THE ISSUE

• Predicting future outcomes — no matter the endeavor — is compelling
  • Big data isolates correlations humans cannot see
  • This leads to predictions based on non-intuitive data relationships

• There is exponentially more data in an observational world to drive insights — most beyond the individual’s knowledge and control

• Assessment is no longer based on whether an activity was covered by a notice, but whether an application is a legal, fair, and just application

• This requires understanding the broad range of risks to individuals and whether risk mitigation is effective
IF ISSUE IS NOT RESOLVED

• A vacuum leads first and foremost to reticence risk
  • Your organization doesn’t use data aggressively enough to drive innovation
  • Behavior at other companies will drive norms impacting whole industries

• Legal and regulatory bright lines that prohibit data creation and use
  • It can happen here

• Lost trust – a killer to brands
WHY IS BIG DATA DIFFERENT?

• Volume, velocity, diversity – blending data robustly
  • More room for processing errors
  • Diverse data sets stretch link to original context
  • To quote Ken Cukier: “more is different”

• Goes beyond testing best intuition – data drives questions
  • Data correlations drive further questions and then insights

• Correlation not causation
  • Is the correlation meaningful
BIG DATA RISKS TO INDIVIDUALS AND SOCIETY

• Data not understood leading to false correlations
• Data understood but the science badly done leading to false correlations
• Correlations correct but users forget it is probability leading to incorrect actions
• Correlations correct but applications are socially unacceptable
• Hunger for insights leads to observation beyond acceptable norms
• Impact assessments must address these
ASSESSMENTS BEGIN WITH FLEXIBILITY IN THE WHITE SPACE BETWEEN THE LINES IN LEGISLATION

• EU law includes legitimate interests based on balancing process
  • It requires a trusted balancing process
• FTC consent orders include requirements for accountability based in part on privacy by design
  • Include information impact assessments
  • Guidance hasn’t defined a balancing process
ASSESSMENT PROCESSES MUST ASSURE PROCESSING IS LEGAL, FAIR, AND JUST

CIVIL LAW DATA PROTECTION
• All processing must have a legal basis
  • Where consent is ineffective one must demonstrate:
    • Fair
    • Just

U.S. COMMON LAW
• Processing must not be illegal
  • But absence of legal prohibition is not enough, must be:
    • Not unfair
    • Just
IAF AND BIG DATA WORK

• 2007 – Current team members began thinking about big data

• 2013 – Two-phase (discovery and application) approach by Abrams, Bruening, and Leta

• 2014 – Challenge to develop an assessment process that would be implementable, could be demonstrated, and could lead to trust

• 2014 – Unified Ethical Frame for Big Data Analysis

• 2015 – Enforcement Study
STARTS WITH DIFFERENTIATING PRIVACY AND DATA PROTECTION

• What do we mean when we say privacy?
  • There are many definitions

• EU legitimate interests paper confronted this issue
  • EU privacy law is about autonomy and family life
  • Data protection assures the full range of fundamental rights

• For purposes of this discussion, we build on EU definitions
  • Privacy relates to the protection of individual autonomy and family life
  • Data protection is foundational for the protection of the full range of individual interests in an information society
KEY GOVERNANCE CONCEPTS

• Data protection assures the full range of individual interests, not just a narrow definition of privacy

• Reticence risk is meaningful, real, and creates harms

• A fair and just assessment is necessary – what is the cost of not processing

• This takes us beyond compliance to ethics
A UNIFIED ETHICAL FRAME FOR BIG DATA

Big Data provides **unprecedented opportunities** to drive innovation in economies, healthcare, public safety, education, transportation, and almost every human endeavor.

Big Data also **creates risk** to both individuals and society unless effective governance is in place.

Governance must be sensitive to the full range of interests. **Ethics must form the basis of decisions** balancing the incentives that come with free enterprise and the breadth of human interests.
1. Corporate commitment to internal policies (codes of conduct) that link to external criteria – data protection law
2. Mechanisms to put those policies into effect, including identifying risk to individuals and mitigating those risks (privacy-by-design)
3. Internal monitoring to assure mechanisms work
4. Individual participation – transparency; consent (where effective)
5. Standing ready to demonstrate to a regulator on request, and remediation where necessary
BIG DATA CODE OF ETHICAL PRACTICE – STRUCTURE

Part A

• Description of the unified ethical framework
• Creates a basis for the assessment guidance

Part B

• Assessment guidance for implementing the code
• Illuminates the key issues that must be considered in making a judgment on whether a Big Data project is fair, responsible and ethical

Part C

• Mechanisms for enforceability
• Demonstrates compliance with code.

Part D

• Contextual assessment questionnaire
• Customized for organizations, industries, mediums.
Big Data requires an assessment against a variety of factors, including existing laws as well as the ethical frames of the stakeholders impacted by the processing.

This requires an assessment of the intersection between the various ethical frames; resulting in a more united ethical frame with positive outcome for stakeholders.

To facilitate assessment, the Foundation adopted five key values that when taken as a whole assist in assessing what is legal, fair, and just.

The unified ethical code makes use of the full range of fundamental rights as recognized by data protection law.
KEY VALUES FOR ASSESSMENT

• Core Values
• How they align with FIPs
• Creation of a framework for assessment, Part B
## VALUES FOR AN ETHICAL FRAME

| Beneficial                      | • Sense of purpose  
|                                | • Define the benefits  
|                                | • Parties who gain benefit  
|                                | • Risk-Benefit balance  
| Progressive                    | • Materially better  
|                                | • Measures  
|                                | • Support innovation  
|                                | • Consider risks  
| Sustainable                    | • Legal basis/permissions  
|                                | • Influence of model(s)  
|                                | • Longevity of insights  
|                                | • Ability to refine/correct  
| Respectful                     | • Data origination/restrictions  
|                                | • Context/Purpose  
|                                | • Interests of all parties  
|                                | • Expectations/Rights  
| Fair                           | • Insights/Applications  
|                                | • Inferences/Discrimination  
|                                | • Labeling/Predestination  
|                                | • Compatibility/Legal Basis  |
PART B – ASSESSMENT FRAMEWORK

• Full project description with clear intents
• Questions to identify issues related to processing and accountability
• Clear description of stakeholders
• A description of intended benefits and possible risks
• A final assessment of fairness based on facts
WHEN SHOULD ASSESSMENT TAKE PLACE

• Discovery
  • Project scoping
  • Discovery processing

• Application
  • Using insights to predict behavior
  • Reviewing for effectiveness
TRANSLATING VALUES INTO PROCESSING STEPS

- Assessment process follows business process map
- Values are reflected throughout process
PART B – SECTIONS

• Characterizing the Project
  • Purpose
  • Sources
  • Preparation
  • Contractual and Legal Conditions
  • Accuracy
  • Insights
  • Outcomes
  • Accountability
  • Stakeholders
PART B – SECTIONS CONTINUED

• Beneficial
  • Benefits
  • Risks/Mitigations
  • Risk/Benefit Analysis

• Fair
ASSESSMENT TOOL AND GOVERNANCE – PART D

• Based on Part B framework
• Customized to an industry or company
• Sits within a broader corporate governance structure
• Is demonstrable to a regulator
MARKETING SUBGROUP CHALLENGE

• Create an assessment tool for marketing
  • Based on the unified ethical frame
  • Would be approachable for organization both facilitating and using big data for digital marketing
  • Would enhance decision making
  • Be describable to stakeholders
  • Demonstrable to regulators
BIG DATA ANALYTICS – IS THE PROJECT FAIR?

Fairness

Project Purpose

Sources ➔ Prep (ETL) ➔ Insights ➔ Application

Legal Obligations ➔ Accountability ➔ Stakeholders ➔ Benefits/Risks/Mitigation

Fairness
# Big Data Analytics – Understanding the Data

## Purpose Interrogation
- Project Overview
- Project Sensitivity
- New or Expanded Insights

## Source Interrogation
- All Sources
- All Elements & Sensitivity
- Origin of Data
- PII Linkability
- Data Structure
- Source Transparency for Use
- Source Accuracy

## Preparation Interrogation
- Consistent Formatting
- Other Synthesizing
- Integration Technique
- Integration Accuracy

## Insights Interrogation
- Expected
- Unexpected
- Use(s)
- Acceptable Accuracy/Use
- Sensitivity/Use
- Useful Life
- Value Over Time

## Application Interrogation
- Better than Current
- Positive Stakeholder Outcomes
- Negative Stakeholder Outcomes
- Public Policy Impact
<table>
<thead>
<tr>
<th>Contractual and Legal Interrogation</th>
<th>Accountability Interrogation</th>
<th>Stakeholder Interrogation</th>
<th>Benefits &amp; Risks Interrogation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source Obligations</td>
<td>Accountable Senior Leader</td>
<td>Source Stakeholders</td>
<td>Benefits for Stakeholders</td>
</tr>
<tr>
<td>Application Obligations</td>
<td>Other Project Leaders</td>
<td>Insight Stakeholders</td>
<td>Risks for Stakeholders</td>
</tr>
<tr>
<td>Security Obligations</td>
<td>All Leaders Comfort with Project</td>
<td>Use Stakeholders</td>
<td>Mitigations for Each Risk</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All Stakeholder Concerns</td>
<td>Residual Risks</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Risk/Benefit Balance</td>
</tr>
</tbody>
</table>
BIG DATA ANALYTICS – DECIDING

<table>
<thead>
<tr>
<th>Fairness Interrogation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unfairness to Individuals</td>
</tr>
<tr>
<td>Issues from Project</td>
</tr>
<tr>
<td>Balance of Residual Risks and Benefits to Individual and Society</td>
</tr>
</tbody>
</table>
INTRODUCTION OF PART D MARKETING ASSESSMENT

• Part D is a customized Part B

• Customized for the Digital Marketing Industry

• Other customizations will be done for sectors, industries, companies

• Major sections are same as Part B
  • Characterizing the Project
  • Beneficial
  • Fair
HYPOTHETICALS

• Break into two groups
• Each group will work through the assessment framework using one of the hypotheticals
• Groups will reconvene to discuss results
HYPOTHETICALS: REVIEW ASSESSMENT REPORTS

• Discuss Results
OVERSIGHT AND ENFORCEMENT

• Conversation with Genie and Marty on domestic and international oversight and enforcement elements
ENCOURAGEMENT FOR ASSESSMENTS

• Positive incentives – freedom to use data for advanced analytics
  • BCR-like certification for big data

• Neutral incentives – market, political, or regulator guidance on best practices
  • Business codes for digital marketing
  • Potential extension of accountability guidance to big data

• Negative incentives – enforcement
  • Codes not followed
  • Bad results from not using suggested best practices
  • Deception about internal policies not followed
OVERSIGHT OF WHAT

• Data compatible or in context
  • Discovery
    • Does it meet plausible tests for being research
  • Compatible for application
    • May follow tradition legal tests

• Is the processing ok
  • Fair and just
  • Legal based on fair and just
  • Explicitly legal
RELEVANCE

• An ethical big data assessment process is the future of today’s information impact assessments

• They form the basis for legitimate interest analysis as required in other jurisdictions

• There are other models – they rest on the foundation of the “Unified Ethical Frame”
TIMING

- The time for planning is now
- The business demand will be sooner than you think
LEVELS OF CERTIFICATION

• Industry codes of conduct endorsed by regulatory structure
• Professional codes of practice – for example for data scientists
• Industry code not endorsed or recognized by regulator
• Internal code of conduct based on external criteria certified by external body
• Internal code of conduct publically announced
CERTIFICATION BY WHOM

- Self-certification
- Self-certification backed by an accountability agent
- Self-certification backed by a regulator
- Certification by an accountability agent
- Certification by a regulator
QUESTIONS

• mabrams@informationaccountability.org
• lgoldstein@informationaccountability.org
• Gbarton@council.bbb.org
• Sheila.Colclasure@acxiom.com