Two Aspects of Data Integrity
Two Aspects of Data Integrity

• How accurate are the data?
  ○ Are they what they represent themselves to be?
  ○ How much noise (imprecision) is in the data?
  ○ Questions of quality and validity

• How ethical was the gathering of the data?
  ○ Was permission granted?
  ○ Is privacy being protected?
  ○ Questions of fairness and propriety
n·teg·ri·ty
inˈtegrədē/
noun
1.1.
the quality of being honest and having strong moral principles; moral uprightness.
"he is known to be a man of integrity"
2.2.
the state of being whole and undivided.
"upholding territorial integrity and national sovereignty"
1. The Data Labeling Initiative

- **Ingredients label metaphor**
- **Collaboration among**
  - DMA (ANA)
  - ARF
  - CIMM
  - IAB
  - Advertisers & adtech companies
- **Unveiled yesterday afternoon**
- **Compliance/transparency**
2. Data Validity Initiative

- **Thermometer metaphor**: Can a cost-effective online survey be used to judge the **accuracy** of digital targeted segments?
  - Partners in ARF Study
    - Lucid and comScore
    - LiveRamp and ODC
  - Other investigators working independently, sharing results with the ARF
    - ODC-Survata
    - LiveRamp-Lucid
    - Lotame
    - Sequent Partners
    - Selective advertisers
Attribute Density: Simple Index of Data Quality

CONCEPT:

• For any given dataset target group
  ○ Survey a sample of members of the dataset group
  ○ Ask industry-standardized question about membership in the claimed target group
  ○ Dataset Attribute Density = % who said “yes” to question
  ○ Normalized Density = DAD indexed to incidence in general population

Tells you how much dataset improved marketers’ odds of reaching target, relative to random
A Proof of Concept Inquiry

• Can a simple, cost-effective method be validated to measure DAD and NAD?

• The Complications
  ○ Best truth sets to use as benchmarks?
  ○ Alternative sampling strategies?
  ○ Differential efficacy for
    • Behaviors vs. attitudes vs. attributes
    • Incidence levels
    • Time windows
  ○ Alternative question formats

• Data collection not complete: very preliminary findings
## Five Segments Tested

<table>
<thead>
<tr>
<th>Segment</th>
<th>Survey Question</th>
<th>Benchmark</th>
<th>Benchmark Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto Intender</td>
<td>Intent to purchase or lease 12 months</td>
<td>MRI</td>
<td>Purchased or lease last 12 months</td>
</tr>
<tr>
<td>Cereal Bar</td>
<td>Total number used last 30 days:</td>
<td>MRI</td>
<td>You or someone in your HH frequently purchase:</td>
</tr>
<tr>
<td>SkyMiles</td>
<td>Which of the following rewards programs:</td>
<td>MRI</td>
<td>Are you a frequent flyer of Delta</td>
</tr>
<tr>
<td>Movie Goer</td>
<td>Which of the following places have you visited twice in last 30 days:</td>
<td>MRI</td>
<td>How often have you visited a movie theater: 2+ month</td>
</tr>
<tr>
<td>Investable Assets</td>
<td>What is the total value of your hh’s security assets including stocks, bonds …</td>
<td>MRI</td>
<td>Please mark the securities and/or … savings plans you own: $50K+</td>
</tr>
</tbody>
</table>
## Segment Sample Sizes

<table>
<thead>
<tr>
<th>Segment</th>
<th>Reference</th>
<th>Segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto Intender</td>
<td>267</td>
<td>384</td>
</tr>
<tr>
<td>Cereal Bar</td>
<td>299</td>
<td>324</td>
</tr>
<tr>
<td>SkyMiles</td>
<td>314</td>
<td>350</td>
</tr>
<tr>
<td>Movie Goer</td>
<td>287</td>
<td>353</td>
</tr>
<tr>
<td>Investable Assets</td>
<td>289</td>
<td>369</td>
</tr>
</tbody>
</table>
Segment Density Compared to Reference and Indexed to Benchmark

* Index Affected by Seasonality
Adding Gender Improves Targeting

Auto Intender

Auto Intender By Gender

<table>
<thead>
<tr>
<th>Gender</th>
<th>Benchmark</th>
<th>Segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>127</td>
<td>164</td>
</tr>
<tr>
<td>Male</td>
<td>139</td>
<td></td>
</tr>
</tbody>
</table>
Tentative Conclusions

• Survey tool appears to be viable, but...

• Choice of reference set is critical
• Commercial reference set may not be a good benchmark
• Targeting more efficient for low-incidence categories
• Behaviors easier to index than attitudes/intentions
• Hybrid targets (demo + behavior) may be better
Validation data still being collected, so all results are very preliminary!

Now we pivot from data quality to data ethics...
Is Privacy Obsolete?

“You have zero privacy...get over it”
Scott McNally, CEO of Sun Microsystems, 1999

“The social norms of privacy have evolved...people [have] really gotten comfortable not only sharing more information and different kinds, but more openly and with more people.”
Mark Zuckerberg, Founder & CEO of Facebook, 2010
Explosive Growth of Data Reducing Privacy

• Search, app, browse and download behaviors
• Social media posts & social network maps
• Email scans
• Location data
• Purchase data
• Biometric and sensor data
• Facial recognition
• Security cameras
• License plate readers
• Stingrays (call intercepts)
• Doppler radar and deep vision
Privacy Backlash

• GDPR
• State laws
  ◦ CCPA in California
  ◦ Vermont, Colorado, Illinois
• FTC Hearings
• Polls show rising anxiety
• Industry self-regulation?
GDPR: What Is It

- 86 page EU regulation governing the protection of EU citizens no matter where they are
- Some Key Features
  - Freely given unambiguous consent to drop cookies
  - Analytics may require second consent
  - Consent for profiling and auto-decisions
  - Data Protection Officer
  - Access to data
  - Ability to Correct and Delete
- Allowable contacts if prior: contract, loyalty program, contest, membership, balanced solicitation with prior purchase
Sticking Points

• What counts as PII?
• Who owns your data?
  ◦ What rules govern the buying and selling of data?
  ◦ What about matching data & building profiles?
• What is “freely given consent”?
  ◦ Are standard T&C documents sufficient?
  ◦ Do consumers understand tradeoffs?
• Can consent given in one use-context extend to another?
2018 ARF Privacy Study Shows

• Consumer willingness to share aggregate, descriptive data with trusted websites, but
  ○ Resistance to sharing data that permit identification in real world (e.g. email, name/address)
  ○ Resistance to sharing sensitive info (medical, financial, govt)
• No change in willingness in exchange for customization
• High bar for PII: low $ value assigned to descriptive data, but view PII seen as priceless
• Very poor comprehension of terms used in T&C forms
People understand the benefit, but they don’t understand the tools

We use various technologies to collect and store information, including cookies, pixel tags, local storage, such as browser web storage or application data caches, databases, and server logs.

We combine our first party data with third party data to serve you ads that are more interesting to you.

Legend: Clear Neither clear nor confusing Confusing
<table>
<thead>
<tr>
<th>Statement</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Concerned about the current privacy practices of most data collectors”</td>
<td>87%</td>
</tr>
<tr>
<td>“I am nervous about burglaries, stalkers or digital/physical harm when sharing my location data”</td>
<td>77%</td>
</tr>
<tr>
<td>“Would be willing to share their location data if they knew there was a financial or other benefit when the data collector sells their data to a third party”</td>
<td>53%</td>
</tr>
</tbody>
</table>
This is especially true for sharing location data when the benefit was related to the APP. For example:

<table>
<thead>
<tr>
<th>Safety in the car, Saving money, Enabling a service</th>
<th>73%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enabling a personalized service</td>
<td>55%</td>
</tr>
<tr>
<td>Personalized advertising</td>
<td>26%</td>
</tr>
</tbody>
</table>
HERE Research

- Strong lack of trust in data collectors, especially location
- One third of consumers are very restrictive of location data for any reason
- Car safety and navigation are the most likely reasons for sharing location
- Consumers do not actively update their location settings
- The greater the transparency, the greater the willingness to share data

• 93% Collecting data is essential for my business

• Once collected, who owns the data?
  ○ 64% My Company
  ○ 27% The Data Subject

• 51% Businesses share data too freely with 3rd parties

*US data only
The ARF’s Member Survey on GDPR

• Slower response than expected
• Possible reasons?
  ○ Province of IT?
  ○ Still a work in progress?
  ○ Lack of regulatory clarity?
  ○ More cautious environment?

• 20 companies responded
  ○ So the usual qualitative caveats apply
Online survey sent out to all ARF member ambassadors
Field Dates: September 11–October 1
N = 20 Completes
65% have European Visitors
25% not sure
One reason for low response was people not really sure what their companies are doing
Leading Steps Members Have Taken

- Consulted Legal Council
- Revised Privacy Policy
- Revised opt in/out
- Requested Cookie Permission
- Use Implies Consent
- DPO or Data Controller
- Altered Targeting
- Transparency & Consent Framework

% Respondents
Which of the Following has GDPR Affected?

- Analytics: 55%
- Marketing: 55%
- Legal: 40%
- IT: 35%
- Digital Ads: 35%
The Present and Future

- Relative to expectation, has GDPR affected your business?
  - Less so: 70%

- Are you likely to take future steps”
  - Yes: 70%
    - Yet there was no consistency nor certainty regarding future steps

- Are you aware of the California Consumer Privacy Act?
  - Yes: 60%

- Do you expect legislation like GDPR or CCPA to be enacted by 2020
  - No: 7%
    - 93% yes or unsure

- Will it change the industry’s use of targeting?
  - Yes: 71%
The Present and Future

• Will future legislation be burdensome for your business?
  ○ Yes: 43%
  ○ No: 43%

• Are current laws sufficient to protect consumers?
  ○ Yes: 7%
  ○ No: 57%
The ARF Code of Conduct

• ARF’s Unique Membership Base
  • Media, Ad Tech, Advertisers, Agencies, Research, Consultants

• What makes our code different?
  • Privacy Commitment requires
    • Research on Use and Understanding of the Terms
  • Chain of Trust Principle
  • Declaration of Definition of PII
  • Disclose
    • Online Behavioral Advertising
    • Automated and AI Based Decision Making

• Covers
  • Withdrawal of Consent
  • Cookie Deletion
  • Location-Based Data
Panelists

- Abby Mehta, Bank of America
- Pete Doe, Clypd
- Michael Schoen, Neustar
- Jon Stewart, Survata
Upcoming ARF Events

Winning at OTT
OTTxSCIENCE 2018
October 18, Universal City, CA

Expand Your Expertise
LEADERSHIP LAB
November 8, New York City

Better Predictions for What Comes Next
FORECASTxSCIENCE 2018
November 14, San Francisco, CA

Build Community and Hone Skills
WOMEN IN ANALYTICS
December 4, New York City